Biochemical and Crystallographic Studies of Superfamily One Helicases

Stuart Peter Griffiths

Molecular Enzymology Laboratory

Cancer Research UK

and

University College London

A thesis submitted for the degree of Doctor of Philosophy at the University of London

2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

To my wife

"Happy is he who gets to know the reasons for things"

Virgil

Acknowledgements

I would like to thank my supervisor, Dale Wigley, for the opportunity of working with him and for all his help during my studies. I really appreciated the fact that his door was always open and that he would drop whatever he was doing when I needed to talk to him. Thank you to everyone in the Molecular Enzymology Laboratory, past and present, whose support, advice and friendship were invaluable toward my PhD. I will miss the "That's a job for google." conversations at lunchtime. Thank you to Robert Court for his supervision in the lab, his patience when I wouldn't listen to him and his brilliant example of how to go about life. Thank you to Jane Sandall for being my personal spell checker and editor, without her help this thesis would be all the poorer. Thank you to my parents for always encouraging my interest in science and for supporting me through my BSc and my PhD. I have enjoyed my time at Cancer Research UK, thank you to all the friends who made it so much fun.

Finally I would like to thank my wife Karen, without whose love, faith and support I would be lost.

Cancer Research UK funded this work.

Declaration

This thesis is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated in the text.

Stuart Peter Griffiths

Abstract

Helicases unwind duplex DNA and are required for a multitude of cellular activities. The majority of Superfamily one (SF1) helicases unwind duplex DNA with 3' – 5' directionality and have been studied in depth. However, there is a set of SF1 helicases with 5' – 3' directionality, the mechanism of which is yet to be understood.

The initial research of this thesis was to determine the role of the 2B domain of the SF1 3' -5' helicase, *Bacillus stearothermophilus* PcrA using modified inteins. However, this work, and work attempting to use the same modified inteins to study the Replication Factor C dependent loading of Proliferating Cell Nuclear Antigen onto DNA, proved unsuccessful.

The crystal structure of the *E.coli* RecBCD complex has been solved previously and gives some insights into the 5'- 3' helicase mechanism of the SF1 helicase RecD. Due to stability problems, isolated *E.coli* RecD protein is an unsuitable target for characterisation; therefore RecD homologues from other species were examined. The RecD homologue from *Deinococcus radiodurans* (*dr*RecD2) was found to be soluble and was chosen for biochemical and structural characterisation.

Mutational studies were carried out to investigate the molecular mechanism by which *dr*RecD2 unwinds duplex DNA. A molecular pin has been identified and is implicated in splitting open duplex DNA as part of the helicase mechanism. In addition to biochemical characterisation, *dr*RecD2 was crystallised with a 5'-tailed duplex DNA substrate. The crystalisation conditions were optimised leading to diffraction data of 3.5Å resolution.

Table of contents

1 INTRODUCTION 15

1.1 Helicases **15**
1.2 Biochemical properties of helicases **16**
1.2.1 DNA binding 16
1.2.2 NTP hydrolysis 17
1.2.3 Rate of translocation 17
1.2.4 Directionality 17
1.2.5 Processivity 18
1.2.6 Step size 19
1.2.7 Oligomeric state 20
1.3 Classification **21**
1.4 The RecA-like fold core domains **22**
1.5 The helicase motifs **24**
1.6 Further classification **26**
1.7 Accessory domains and proteins **27**
1.8 Models of helicase activity **27**
1.8.1 Active and passive mechanisms 27
1.8.2 Superfamily 1 helicases 29
1.8.3 The inchworm model 30
1.8.4 The active rolling model 32
1.9 Superfamily 1A **35**
1.10 Superfamily 1B **35**
1.10.1 *E.coli* RecD 35
1.10.2 Opposite directionality of Superfamily1Bα helicases compared Superfamily1Aα helicases 37
1.10.3 Type 1 and 2 RecD helicases 40
1.10.4 *Deinococcus radiodurans* RecD2 43
1.10.5 Eukaryotic Superfamily 1B helicases 43
1.10.5.1 The Pif1-like helicases 43
1.10.5.2 The human Upf1 helicase 44
1.10.5.3 The BRCA1-associated C-terminal helicase 1 (BACH1) 45
1.11 Thesis outline **45**

2 MATERIALS AND METHODS 47

2.1 Standard materials and methods **47**
2.1.1 Standard methods 47
2.1.2 Standard materials 47
2.2 Microbiological methods **47**
2.2.1 *E.coli* strains 47
2.3 DNA methods **49**
2.3.1 Oligonucleotide purification 49
2.3.2 Annealing oligonucleotides 49
2.3.3 Cloning 49
2.3.4 Site-directed mutagenesis 52
2.4 Protein methods **53**
2.4.1 Protein expression 53
2.4.1.1 Protein expression trials 53
2.4.1.2 Protein expression details 53
2.4.2 Protein purification 55
2.4.2.1 *Deinococcus radiodurans* RecD2 small scale purification 55
2.4.2.2 *Deinococcus radiodurans* RecD2 and RecD2 mutants large scale purification 55
2.4.2.3 Seleno-methionine *Deinococcus radiodurans* RecD2 56
2.4.2.4 *E.coli* RecD 56
2.4.2.5 Small scale purification of proteins fused to either *Mxe*GryAi or *Mth*RIR1i 57
2.4.2.6 Large scale purification of proteins fused to either *Mxe*GryAi or *Mth*RIR1i 57
2.4.2.7 Large scale purification of proteins fused to *Ssp*DNABi 58
2.4.3 Ligation of intein prepared polypeptides 59
2.4.4 Protein/protein complex analysis by gel-filtration chromatography 59
2.4.5 Dynamic light scattering 59

2.4.6 Identification of proteins by mass spectrometry 60
2.5 Biochemical assay methods 60
2.5.1 DNA substrates 60
2.5.2 Radio labelling of DNA 61
2.5.3 ATP-hydrolysis assay 62
2.5.3.1 Experimental setup 62
2.5.4 DNA-binding assay (gel-mobility shift assay) 64
2.5.4.1 Experimental setup 64
2.5.4.2 Theory relating to the calculation of K_d 65
2.5.5 Isothermal calorimetry (ITC) 66
2.5.5.1 Sample preparation 66
2.5.5.2 Experimental setup 66
2.5.5.3 Experimental parameters 68
2.5.6 Helicase assay (strand-displacement assay) 68
2.5.7 Processivity helicase assay 69
2.6 X-ray crystallography methods 70
2.6.1 DNA substrate S1 70
2.6.2 Crystallisation 71
2.6.2.1 Micro-seeding 72
2.6.3 Data collection 72
2.6.3.1 Room temperature 72
2.6.3.2 100 °K 72
2.6.4 Data processing 73
2.6.4.1 Integration and scaling 73
2.6.4.2 Phase determination 74
2.7 Buffer formulations 75
2.7.1 Cell culture 75
2.7.2 Gel electrophoresis 76
2.7.2.1 Loading buffers 76
2.7.2.2 Running buffers 76
2.7.2.3 Staining/destaining 76
2.7.3 Protein purification 77

3 AN ATTEMPT TO MAKE *BACILLUS STEAROTHERMOPHILUS* PCRA MUTANTS AND AN *ARCHAEOGLOBUS FULGIDUS* PCNA DIMER USING THE IMPACT-TWIN INTEIN SYSTEM 78

3.1 Introduction 78
3.1.1 Summary 78
3.1.2 Inteins 78
3.2 An attempt to make full length *Bacillus stearothermophilus* PcrA from two halves 84
3.2.1 Introduction 84
3.2.2 Aims and strategy 85
3.2.3 Cloning and expression of *bs*PcrA1-*Mxe*GyrAi and *bs*PcrA1-*Mth*RIR1i 88
3.2.4 Purification of *bs*PcrA1 from *bs*PcrA1-*Mxe*GyrAi 90
3.2.5 Cloning and expression of *Ssp*DnaBi-*bs*PcrA2 92
3.2.6 Purification of *bs*PcrA2 from *Ssp*DnaB-*bs*PcrA2 92
3.2.7 Ligation of *bs*PcrA1 and *bs*PcrA2 to form *bs*PcrA 93
3.2.8 Other *bs*PcrA N-terminal intein fusion proteins 95
3.3 An attempt to make an *Archaeoglobus fulgidus* PCNA dimer 98
3.3.1 Introduction 98
3.3.2 Cloning and expression of *af*PCNA-*Mxe*GyrAi and *af*PCNA-*Mth*RIR1i 104
3.3.3 Purification of *af*PCNA from *af*PCNA-*Mxe*GyrAi and *af*PCNA-*Mth*RIR1i 105
3.3.4 Cloning and expression of *Ssp*DnaBi- *af*PCNA 107
Discussion 109
3.3.5 *Bacillus stearothermophilus* PcrA intein work 109
3.3.6 *Archaeoglobus fulgidus* PCNA intein work 111

4 BIOCHEMICAL ANALYSIS OF *DEINOCOCCUS RADIODURANS* RECD2 113

4.1 Introduction 113
4.1.1 Summary 113
4.1.2 The Superfamily 1Bα RecD helicase 113
4.2 Identification and purification of *Deinococcus radiodurans* RecD2 115
4.2.1 *E.coli* RecD1 115

4.2.2 Identification of *Deinococcus radiodurans* RecD2 116
4.2.3 Purification of *Deinococcus radiodurans* RecD2 120
4.2.4 Possible complex formation between *Deinococcus radiodurans* RecD2 and *E.coli* RecBC 125
4.3 Biochemical analysis of wild-type *Deinococcus radiodurans* RecD2 128
4.3.1 ATP hydrolysis activity 128
4.3.2 DNA binding activity 130
4.3.2.1 Isothermal calorimetry (ITC) 130
4.3.2.2 DNA binding assays 138
4.3.3 Helicase activity 139
4.4 Biochemical analysis of *Deinococcus radiodurans* RecD2 mutants 141
4.4.1 A new gene for the creation of *Deinococcus radiodurans* RecD2 mutants 141
4.4.2 Motif A 144
4.4.2.1 Introduction 144
4.4.2.2 Biochemical analysis of *dr*RecD2^{F635A} 147
4.4.3 The N-terminal region 149
4.4.3.1 Introduction 149
4.4.3.2 Biochemical analysis of *dr*RecD$^{\Delta NT}$ 151
4.4.3.3 Biochemical analysis of *dr*RecD2^{NT} 158
4.4.4 The 1B domain 161
4.4.4.1 Introduction 161
4.4.4.2 Biochemical analysis of *dr*RecD2$^{\Delta loop}$ 164
4.5 Discussion 169
4.5.1 Identification and purification of *Deinococcus radiodurans* RecD2 169
4.5.2 Biochemical analysis of wild-type *Deinococcus radiodurans* RecD2 171
4.5.3 The role of phenylalanine 635 of motif A 175
4.5.4 The role of the N-terminal region 176
4.5.5 The role of the 1B domain 179
4.5.6 Conclusions 181
4.6 Future Work 182

5 CRYSTALLISATION OF *DEINOCOCCUS RADIODURANS* RECD2 185

5.1 Introduction 185
5.1.1 Summary 185
5.1.2 Current structural understanding of Superfamily 1Bα helicases 185
5.2 Crystallisation of *Deinococcus radiodurans* RecD2 186
5.2.1 Original screening 186
5.2.2 Optimisation 189
5.2.3 Growing discrete crystals 189
5.2.4 Optimisation of discrete crystals 194
5.2.5 Data collection 197
5.2.5.1 Seleno-methionine *dr*RecD2 197
5.2.5.2 Heavy atom derivatives 198
5.2.5.3 Iodouridine DNA derivative 202
5.2.6 Data processing 202
5.2.7 Phase determination 205
5.2.7.1 Native crystals 205
5.2.7.2 Molecular replacement 205
5.2.7.3 Heavy atom derivatives 206
5.3 Discussion 209
5.4 Future work 211

6 BIBLIOGRAPHY 213

7 APPENDIX 223

7.1 Sequence comparison of RecD 223

Table of figures

Figure 1.1. A generic helicase reaction 15
Figure 1.2. Helicase directionality 18
Figure 1.3. The RecA-like fold core domains common to all helicase Superfamilies 23
Figure 1.4. The Superfamily 1 helicase motifs 24
Figure 1.5. Further classification of helicases based upon biochemical properties 26
Figure 1.6. The inchworm model of helicase action 30
Figure 1.7. The cooperative inchworm model 32
Figure 1.8. The DNA bound states of a Rep dimer 33
Figure 1.9. The active rolling model for the *E.coli* Rep helicase 33
Figure 1.10. The structure of *E.coli* RecBCD complex 36
Figure 1.11. Models of how the 3' – 5' directionality of a Superfamily 1Aα helicase could be reversed to 5' – 3' 38
Figure 1.12. The orientation of DNA binding by Superfamily 1Aα and Superfamily 1Bα helicases. 39
Figure 1.13. An un-rooted phylogenetic tree of the RecD1 and RecD2 proteins 42
Figure 2.1. The cloning of *dr*RecD2$^{\Delta loop}$ 52
Figure 2.2. The 5'-tailed duplex substrate 60
Figure 2.3. 40bp 5'-tailed duplex substrate 61
Figure 2.4. The dsDNA substrate 61
Figure 2.5. The coupled ATP-hydrolysis assay 62
Figure 2.6. DNA S1 70
Figure 3.1. Cloning P1 and P2 into pTWIN1/pTWIN2 80
Figure 3.2. Strategy to purify and ligate P1 and P2 protein fragments to make the P1-P2 fusion protein 81
Figure 3.3. Peptide bond formation between P1 and P2 82
Figure 3.4. The structure of *bs*PcrA 84
Figure 3.5. Strategy for creating *bs*PcrA deletion mutants 86
Figure 3.6. Crystal structure of *bs*PcrA 87
Figure 3.7. Expression of *bs*PcrA1-*Mxe*GryAi and *bs*PcrA1-*Mth*RIR1i 89
Figure 3.8. Small scale purification tests of *bs*PcrA1 91
Figure 3.9. Purification of *bs*PcrA1 92
Figure 3.10. Purification of *bs*PcrA2 93
Figure 3.11. Ligation of *bs*PcrA1 and *bs*PcrA2 to form full length *bs*PcrA 95
Figure 3.12. Potential intein fusion points within *bs*PcrA 96
Figure 3.13. The crystal structure of *af*PCNA 98
Figure 3.14. Possible ways for a PCNA trimer to be loaded onto DNA 100
Figure 3.15. Fusing two PCNA monomers 102
Figure 3.16. Strategy for creating an *af*PCNA dimer 103
Figure 3.17. Small-scale purification tests of *af*PCNA fused to either *Mxe*GryAi or *Mth*RIR1i 106
Figure 3.18. Purification of *af*PCNA from *af*PCNA-*Mth*RIR1i 107
Figure 3.19. Expression of *Ssp*DnaBi-*af*PCNA 108
Figure 4.1. Purified *E.coli* RecD1^{his} 115
Figure 4.2. Initial expression trial of *Deinococcus radiodurans* RecD2 118
Figure 4.3. Expression trial of *dr*RecD2 with a C-terminal histidine tag 119
Figure 4.4. Purification trials of *dr*RecD2 120
Figure 4.5. Purification of *dr*RecD2 using nickel chelating and heparin columns 122
Figure 4.6. Purification of *dr*RecD2 , MonoQ column 123
Figure 4.7. Gel filtration and dynamic light scattering analysis of *dr*RecD2 125
Figure 4.8. Complex formation of *dr*RecD2 and *E.coli* RecBC 126
Figure 4.9. Co-elution of *E.coli* RecBC and *dr*RecD2$^{\Delta NT}$ from a MonoQ column 127
Figure 4.10. Complex formation of *dr*RecD2$^{\Delta NT}$ and *E.coli* RecBC 127
Figure 4.11. ATP hydrolysis activity of *dr*RecD2 129
Figure 4.12. Determination of K_{DNA2} 129
Figure 4.13. ITC analysis of *dr*RecD2 with various pdT substrates 131
Figure 4.14. Binding events that are indistinguishable by ITC 133
Figure 4.15. Two equilibrium constants in one ITC experiment 135
Figure 4.16. ITC analysis of *dr*RecD2 with pdT20-24 136
Figure 4.17. DNA binding activity of *dr*RecD2 139
Figure 4.18. Helicase activity of *dr*RecD2 140
Figure 4.19. Expression of *dr*RecD2 protein from the *dr*RecD2$_{new}$ gene 142

Figure 4.20. ATP hydrolysis and helicase activity of *dr*RecD2 protein from *dr*RecD2$_{new}$ gene 143
Figure 4.21. Helicase motifs 1a and 3 play a role in the ssDNA translocation by *bs*PcrA. 144
Figure 4.22. Translocation of ssDNA by domain 1A and domain 2A of *bs*PcrA 145
Figure 4.23. A sequence comparison of the helicase motifs 1a and 3 in RecD and *bs*PcrA 145
Figure 4.24. T he ssDNA-binding pockets of *bs*PcrA and *E.coli* RecD1 146
Figure 4.25. Motif A 147
Figure 4.26. Purified *dr*RecD2^{F635A} 148
Figure 4.27. ATP hydrolysis and helicase activity of *dr*RecD2^{F635A} 148
Figure 4.28. A comparison of the N-terminal domains of RecD2 and RecD1 enzymes 150
Figure 4.29. Purity of *dr*RecD2$^{\Delta NT}$ 151
Figure 4.30. ATP hydrolysis activity of *dr*RecD2$^{\Delta NT}$ 152
Figure 4.31. DNA binding activity of *dr*RecD2$^{\Delta NT}$ 153
Figure 4.32. Isothermal calorimetry of *dr*RecD2$^{\Delta NT}$ and pdT15 154
Figure 4.33. Helicase activity of *dr*RecD2$^{\Delta NT}$ 155
Figure 4.34. Processivity assay of *dr*RecD2$^{\Delta NT}$ and *dr*RecD2 157
Figure 4.35. Purity of *dr*RecD2^{NT} 158
Figure 4.36. DNA binding activity of *dr*RecD2^{NT} 159
Figure 4.37. Gel-filtration analysis of *dr*RecD2^{NT} 159
Figure 4.38. The 1B domain loop of RecD 161
Figure 4.39. A comparison of RecD1 and Upf1 163
Figure 4.40. Deleting the 1B loop of *E.coli* RecD1 164
Figure 4.41. Analysis of the purity of *dr*RecD2$^{\Delta loop}$ 164
Figure 4.42. ATP hydrolysis activity of *dr*RecD2$^{\Delta loop}$ 165
Figure 4.43. DNA binding activity of *dr*RecD2$^{\Delta loop}$ 166
Figure 4.44. Helicase activity of *dr*RecD2$^{\Delta loop}$ 167
Figure 4.45. Helicase assay varying the concentration of *dr*RecD2$^{\Delta loop}$ 168
Figure 5.1. DNA S1 187
Figure 5.2. Complex formation of *dr*RecD2 and DNA substrate 1 188
Figure 5.3. Nucleation and crystal growth 190
Figure 5.4. Discrete crystals of *dr*RecD2/DNA S1 complex 192
Figure 5.5. X-ray diffraction pattern of a discrete crystal of *dr*RecD2/DNA substrate 1 complex 193
Figure 5.6. Crystals grown in PIPES, sodium cacodylate or imizadole 195
Figure 5.7. Room temperature X-ray diffraction of a discrete crystal of *dr*RecD2/DNA S1 complex 195
Figure 5.8. Crystals grown with varying lengths of the 5' ssDNA tail 196
Figure 5.9. Fluorescence wavelength scan of seleno-methionine *dr*RecD2/DNA S1 crystal 198
Figure 5.10. X-ray diffraction of a crystal of *dr*RecD2/DNA S1 complex soaked in ethyl mercuric phosphate 201
Figure 5.11. Iodouridine-substituted DNA S1 202
Figure 7.1. Sequence comparison of RecD 224

Table of tables

Table 1.1. Summary of the presence of the Add AB, RecBCD1 and RecD2 genes in bacterial genomes 41
Table 2.1. Summary of *E.coli* expression strains 48
Table 2.2. Summary of helper expression plasmids 48
Table 2.3. Constructs of this thesis 51
Table 2.4. Protein expression parameters 54
Table 2.5. Experimental parameters for the ITC experiments 68
Table 2.6. Crystallisation screens 71
Table 3.1. Definitions of *bs*PcrA domains 88
Table 3.2. The *bs*PcrA N-terminal fragments chosen to be fused via C-terminal to *Mxe*GryAi 97
Table 3.3. Expression trials for *bs*PcrA N-terminal fragment *Mxe*GyrAi fusion proteins 97
Table 3.4. Summary of expression trials of *af*PCNA-*Mxe*GyrAi and *af*PCNA-*Mth*RIR1i 105
Table 3.5 Expression trial conditions of *Ssp*DnaBi-*af*PCNA 108
Table 4.1. Screening for a soluble RecD helicase 117
Table 4.2. Summary of ITC analysis of *dr*RecD2 binding to various pdT substrates 132
Table 4.3. Chi^2/DoF of fitting one-binding-event or two-binding-event models to the ITC data of *dr*RecD2 binding to pdT20-24 137
Table 4.4. Summary of ITC analysis of *dr*RecD2 with pdT20-25 137
Table 4.5. Summary of ATP hydrolysis and helicase activity of *dr*RecD2 143
Table 4.6. Summary of ATP hydrolysis and helicase activity of *dr*RecD2^{F635A} 149
Table 4.7. Summary of ATP hydrolysis activity of *dr*RecD2$^{\Delta NT}$ 152
Table 4.8. Summary of DNA binding activity of *dr*RecD2$^{\Delta NT}$ 153
Table 4.9. Summary of ITC results of *dr*RecD2$^{\Delta NT}$ 154
Table 4.10. Summary of helicase activity of *dr*RecD2$^{\Delta NT}$ 156
Table 4.11. Summary of ATP hydrolysis activity of *dr*RecD2$^{\Delta loop}$ 165
Table 4.12. Summary of the DNA binding activity of *dr*RecD2$^{\Delta loop}$ 167
Table 5.1. Conditions tested to grow single *dr*RecD2/DNA S1 crystals 191
Table 5.2. Conditions tested in order to grow single *dr*RecD2/DNA S1 crystals by micro-seeding . 192
Table 5.3. Changes in buffer 194
Table 5.4. Summary of heavy atom screening 200
Table 5.5. Summary of datasets from crystals with a short c-axis, processed in MOSFLM 203
Table 5.6. Summary of datasets from crystals with a long c-axis, processed in XDS 204
Table 5.7. Matthews coefficient of native datasets 205
Table 5.8. Summary of R_{iso} between datasets 207

Abbreviations

Standard abbreviations are used for amino acids, elements, nucleotides, protein names and units.

[]	concentration (molar)
α	alpha
β	beta
γ	gamma
ε	extinction coefficient
A_x	absorbance of light of x nM wavelength
aa	amino acid
ADP	adenosine diphosphate
AMPPNP	5'-adenylyl-beta,gamma-imidodiphosphate
AU	absorbance units
ATP	adenosine triphosphate
b	base
bp	base pair
BSA	Bovine serum albumin
BME	β mecaptoethanol
C-terminus/terminal	carboxyl terminus/terminal
Da	Dalton
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
ssDNA	single stranded DNA
DTT	di-thiothreitol
E.coli	*Escheria coli*

EDTA	ethylenediaminetetraacetic acid
HEPES	N-2-hydroxyethylpiperzine-N'-2-ethanesulphonic acid
IPTG	isopropyl-β-D-thiogalctopyranoside
KDa	kilo Dalton
K_a	association constant
K_d	dissociation constant
LB	Luria broth
M	molar
MES	2-(N-morpholino)-ethane sulfonic acid
MESNA	2-mercaptoethanesulfonic acid
MPD	2-methyl-2,4-pentanediol
Mw	molecular weight
N-terminus/terminal	amino terminus/terminal
NAD^+	nicotinamide adenine dinucleotide (oxidised form)
NADH	nicotinamide adenine dinucleotide (reduced form)
nt	nucleotide
OD_{600}	optical density at 600 nM
PCR	polymerase chain reaction
PEG	polyethylene glycol
PIPES	piperazine-1,4-bis(2-ethanesulfonic acid)
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
PAGE	polyacrylamide gel electrophoresis
Tris	tris-(hydroxymethyl)-aminomethane
U	units of protein
UV	ultra violet

v/v	volume per volume
w/v	weight per volume

1 Introduction

The aim of this thesis is to further the understanding of how Superfamily 1 (SF1) helicases unwind duplex nucleic acid, with particular attention to the Superfamily 1B (SF1B) helicases.

1.1 Helicases

A helicase is an enzyme that couples ATP hydrolysis to the unwinding of double-stranded nucleic acid into its component strands. All helicases share fundamental biochemical activities, including non-sequence-specific nucleic acid binding, nucleic acid-dependent NTP hydrolysis (usually of ATP), and duplex nucleic acid separation (see Figure 1.1).

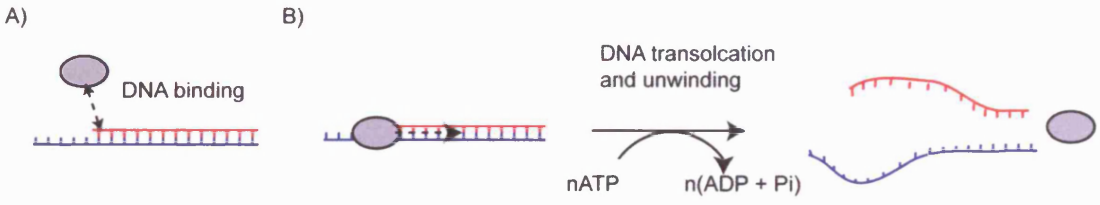

Figure 1.1. A generic helicase reaction

A) The DNA helicase (grey) binds to the DNA substrate (red and blue).
B) DNA translocation and DNA unwinding powered by ATP hydrolysis separates the strands of the duplex DNA.

The first *in vitro* experiments showing a protein to have helicase activity were published in 1976 (Abdel-Monem and Hoffmann-Berling 1976) and the details of the helicase mechanism are still to be fully understood. The ability to unwind DNA means helicases have a significant role to play in almost every cellular process involving nucleic acids. These processes include DNA replication and repair, transcription, translation, ribosome synthesis, RNA maturation and splicing, nuclear

export and conjugal DNA transfer (Tuteja and Tuteja 2004). Helicases are also targets for anti-cancer therapies (Gupta and Brosh 2007) and anti-viral therapies (Xi 2007).

Helicases are a sub-group of translocases, which are enzymes that couple ATP hydrolysis to directional movement on single or double-stranded nucleic acid. A helicase is a translocase with the ability to unwind double-stranded nucleic acid. There are several different ways in which the ability to unwind duplex nucleic acid is conferred to a translocase and these are discussed later.

Helicase activity combines the activities of nucleic acid translocation with double-stranded nucleic acid destabilisation. Both of these activities depend upon other sub-activities. Translocation of single stranded nucleic acid by a helicase depends upon non-sequence-specific nucleic acid binding and ATP hydrolysis activities of the helicase (Lohman and Bjornson 1996).

1.2 Biochemical properties of helicases

There are several biochemical properties of helicases to consider concerning how helicases unwind DNA.

1.2.1 DNA binding

If a helicase either binds nucleic acid too tightly or not at all, translocation cannot occur, i.e. the helicase cannot bind to a substrate or it gets stuck in one place. However, if the nucleic acid binding activity can be modulated between a weak and tight hold on the nucleic acid then this provides an opportunity for single-stranded translocation. The modulation of just one binding site would not be enough to

provide a mechanism of helicase translocation; two DNA binding sites are required (Lohman and Bjornson 1996).

1.2.2 NTP hydrolysis

Energy is required 1) in order to modulate the changing affinity of the helicase for nucleic acid, 2) to overcome the thermodynamically unfavourable reaction of splitting duplex DNA and 3) to confer other structural conformational changes specific to individual helicases. This energy is provided by the nucleic acid-dependent hydrolysis of NTPs by the helicase (Hall and Matson 1999).

1.2.3 Rate of translocation

The rate of translocation is the rate at which a helicase translocates along nucleic acid. This varies immensely from a few base pairs per second as demonstrated by the *Escheria coli* (*E.coli)* Rep helicase (Bjornson *et al.* 1994) to approximately a thousand base pairs per second as demonstrated by the *E.coli* RecBCD complex (Bianco and Kowalczykowski 2000; Roman and Kowalczykowski 1989b). The rate of translocation is usually proportional to the rate of ATP hydrolysis by the helicase.

1.2.4 Directionality

The directionality of a helicase is the direction the helicase translocates along nucleic acid in terms of the asymmetric properties of nucleic acid. Nucleic acid is an asymmetric molecule with a 3' end and a 5' end; helicases that translocate along single-stranded nucleic acid possess a directionality, in that they either move toward the 5' end (a 3' – 5' translocase), or toward the 3' end (a 5' – 3' translocase) (Lohman and Bjornson 1996) (see Figure 1.2).

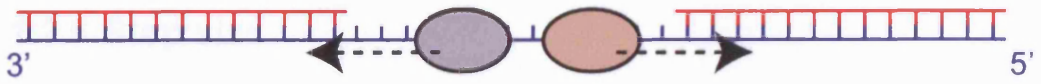

Figure 1.2. Helicase directionality

A schematic showing the direction of translocation along single-stranded nucleic acid by a 5' – 3' helicase (grey) and by a 3' – 5' helicase (brown).

Single-stranded nucleic acid translocases tend to posses an intrinsic directionality regardless of how they are loaded onto the substrate. The situation is more complex for helicases that translocate duplex nucleic acid. Duplex nucleic acid is a bipolar molecule with two strands running anti parallel to one another, which confers the duplex with an intrinsic symmetry. The helicase will translocate along one of the strands with a directionality, which means the directionality is determined by the way the helicase is loaded onto the substrate (Singleton *et al.* 2007).

1.2.5 Processivity

The processivity of a helicase is the number of base pairs unwound before the helicase dissociates from the nucleic acid substrate. Helicases are not distributive enzymes, meaning they do not undergo one round of catalysis and then release the products before rebinding new substrates for the next reaction. Helicases are processive enzymes, meaning they catalyze multiple cycles before releasing the products, i.e. multiple productive cycles of DNA separation and translocation occur for a single nucleic acid binding event before the nucleic acid is released. The processivity of a helicase is dependent on the rate of translocation along, and the rate of dissociation from, the nucleic acid substrate. The *E.coli* RecBCD complex is the most processive helicase unwinding up to 42 kilo bases without dissociating from the DNA substrate (Bianco *et al.* 2001).

1.2.6 Step size

There are various different ideas as to what a step size of a helicase is. The step size is defined here as the number of base pairs that are translocated or unwound in one catalytic cycle (i.e. per ATP hydrolysis reaction). The free energy of ATP hydrolysis (ΔG = -13 kcal/mol (Simmons and Hill 1976)) and an average free energy for base pair separation (ΔG = 1 – 1.5 kcal/mol/bp (Record *et al.* 1981)) should allow the calculation of a theoretical maximum step size. However, these values are highly variable and depend on many factors, for example, the DNA topology. Estimates for a theoretical limit range from 2-4 bp/ATP to 9-12 bp/ATP (Lohman and Bjornson 1996; Roman and Kowalczykowski 1989a). However, these limits assume all the energy from hydrolysis is used to unwind DNA and only a thermodynamically perfect helicase could reach the theoretical limit.

It is very hard to measure a step size defined as the number of bases translocated/unwound per ATP hydrolysised experimentally. The obvious way of measuring a step size is to measure the amount of ATP required to unwind a known number of base pairs. This is appears simple, but there are problems. Uncoupled ATP hydrolysis can occur during nucleic acid dependent ATP hydrolysis, and an unprocessive helicases may dissociate from the nucleic acid allowing strand re-annealing and therefore unproductive helicase activity occurs. These cannot be quantified in the calculation used to analyse the results, and so the step sizes determined will be lower then the true value.

Ali and Lohman were the first to use the approach of investigating pre-steady state kinetics of strand displacement by a helicase using quenched flow apparatus and

applied it to the study of the *E.coli* UvrD helicase (Ali and Lohman 1997). They used 3'-tailed duplex DNA of varying duplex DNA lengths and observed a lag phase before the DNA was unwound suggesting the presence of intermediates on the way to unwinding the all of the DNA. This is thought be an indication of a series of kinetic steps of equal rate, which unwind the substrates. It is important to note that this does not give information relating to the number of ATP hydrolysis events required for each kinetic step. Instead it gives a kinetic step size, which is related to the rate-limiting step of DNA unwinding. The kinetic step could be related to the ATP hydrolysis, but not necessarily.

The structures of SF1 helicases show a clear indication that the step size will be one base pair per ATP hydrolysed, *bs*PcrA and *E.coli* UvrD are the best examples of this (Lee and Yang 2006; Velankar *et al.* 1999). The largest experimentally determined step size in terms of single stranded gaps in duplex DNA that can be passed over with out dissociation of the helicase is 23 nucleotides *E.coli* RecBCD (Bianco and Kowalczykowski 2000). However, the same study estimates that only around 1 base pair is unwound per ATP hydrolysed (Bianco and Kowalczykowski 2000). The next largest example comes from the *E.coli* TraI helicase, which is shown to have a kinetic step size of around 6 to 8 bp (Sikora *et al.* 2006).

1.2.7 Oligomeric state

The oligomeric state of a helicase refers to the number of identical monomers within one active helicase complex. The Superfamily 3, Superfamily 4, Superfamily 5 and Superfamily 6 (SF6) helicases form hexameric rings complexes (Singleton *et al.* 2007). Superfamily 1 and Superfamily 2 helicases can work as monomers *in vitro* (Cheng *et al.* 2007, Singleton, 2006; Morris *et al.* 2001; Nanduri *et al.* 2002;

Velankar *et al.* 1999), but the oligomeric state *in vivo* is unclear and they could form dimers (see section 1.8.4).

1.3 Classification

The classification of helicases was, until recently, based entirely on primary structure analysis as described by Gorbalenya and Koonin (Gorbalenya and Koonin 1993). Gorbalenya and Koonin describe three Superfamilies of helicases, Superfamily 1 (SF1), Superfamily 2 (SF2) and Superfamily 3 (SF3), and two families of, what were then, 'putative helicases' that are now referred to as Superfamily 4 (SF4) (DnaB-like helicases) and Superfamily 5 (SF5) (Rho/V- F-ATPase-like helicases) (Gorbalenya and Koonin 1993). The helicase Superfamily 6 (SF6) has recently been added (Singleton *et al.* 2007) and contains AAA+ (ATPases Associated with various cellular Activities (Erzberger and Berger 2006))-like helicases. The classification of helicases allowed the identification of seven conserved sequence motifs, which are described later (Gorbalenya and Koonin 1993; Gorbalenya *et al.* 1989; Gorbalenya *et al.* 1990). Several new motifs have been identified some of which are only found in the helicases of one Superfamily or a subset of helicases within a Superfamily, for example motif 4a, Q-motif, TRG, and TxGx (Korolev *et al.* 1998; Mahdi *et al.* 2003; Pause and Sonenberg 1992; Tanner *et al.* 2003).

Some proteins do not unwind duplex DNA *in vitro*, but could be classified as helicases based upon primary structure analysis. For example, the *Saccharomyces cerevisiae* (*sc*) PRP2 protein (Kim *et al.* 1992) and the *sc*RAD5 protein (Johnson *et al.* 1994) do not possess helicase activity despite containing the seven helicase motifs and displaying nucleic acid-dependent NTP hydrolysis activity. Indeed, the

helicase motifs may not necessarily be indicative of helicase activity, for example the hsdR subunit of type I restriction endonucleases contains all the motifs characteristic of a Superfamily 2 helicase (Gorbalenya and Koonin 1991). The best characterised set of proteins with the helicase motifs that are not helicases are the chromatin remodelling enzymes. These proteins are part of the Snf2-like family and can be placed with the Superfamily 2 helicases (Lusser and Kadonaga 2003).

The idea of these motifs being 'helicase motifs' may therefore be restrictive as well as being somewhat misleading to those not familiar with the field. However, in all cases to date the helicase motifs are indicative of the structural RecA-like fold 'core domains' of helicases/translocases.

1.4 The RecA-like fold core domains

Helicases of all Superfamilies contain RecA-like fold structural core domains, also referred to as the motor domains (see Figure 1.3).

Figure 1.3. The RecA-like fold core domains common to all helicase Superfamilies

A) i) The *E.coli* RecA fold. ii) A schematic of the fold (based upon a figure by Caruthers *et al.* (Caruthers and McKay 2002). B) Examples of the RecA-like fold core domains of different Superfamily helicases. Where the helicase contains two core domains in the same polypeptide one is coloured red the other green. Two of the six core domains are shown for the ring helicases, except for *Tm*RuvB because the structure of the biological ring has yet to be determined. i) *Bacillus stearothermophilus* PcrA (Velankar *et al.* 1999). ii) *Thermatoga maritima* RecG (Singleton *et al.* 2001). iii) Bovine papillomavirus E1 (Enemark and Joshua-Tor 2006). iv) *Thermus aquaticus* DnaB (Bailey *et al.* 2007). v) *E.coli* Rho (Skordalakes and Berger 2003). vi) *Thermotoga maritima* RuvB (Putnam *et al.* 2001).

These core domains couple NTP binding and hydrolysis to conformational changes within the protein associated with translocation (Ye *et al.* 2004).

1.5 The helicase motifs

For the SF1 helicases, which contain the two core RecA-like domains in one polypeptide, seven helicase motifs were originally described (Gorbalenya and Koonin 1993). For the purposes of this thesis, the two core domains are classified as 1A and 2A, after *Bacillus stearothermophilus* (*bs)* PcrA (Subramanya *et al.* 1996), depending on which helicase motifs they contain (see Figure 1.4). Domain 1A contains motifs 1, 1a, 2, and 3 and is N-terminal to domain 2A, which contains motifs 4, 5 and 6 and is C-terminal to domain 1A.

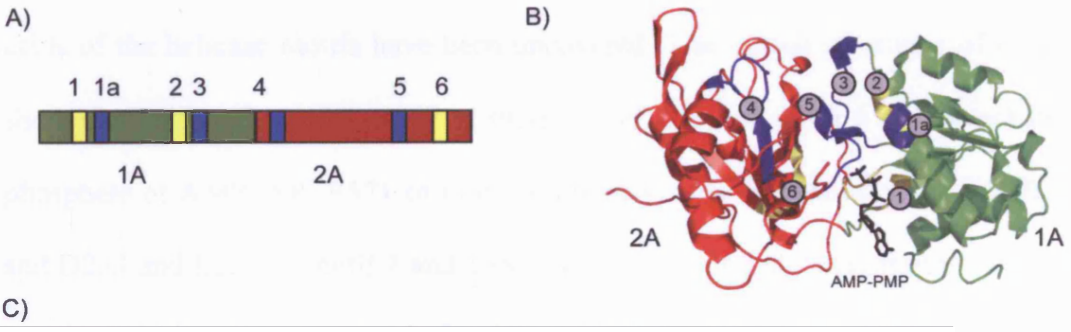

Figure 1.4. The Superfamily 1 helicase motifs

A) A schematic of the *bs*PcrA gene that encodes for the RecA-like core domains with the locations of the helicase motifs indicated. B) The structure of the RecA-like fold core domains of *bs*PcrA and the locations of the helicase motifs. C) The conserved amino acid sequence of helicase motifs of Superfamily 1 helicases (Gorbalenya *et al.* 1989). Capital letters are the single letter abbreviation and for amino acids present in at least 75% of known sequences. + represents a hydrophobic residue, o represents a hydrophilic residue, and x is any residue.

Motifs common to all helicases of all Superfamilies are coloured yellow, motifs coloured blue are specific to SF1 (and SF2) helicases. The 1A domain is coloured green and the 2A domain is coloured red (this colour scheme is the same in all subsequent figures). The 1B and 2B domains of *bs*PcrA are omitted for clarity.

The SF2 helicases also contain the seven helicase motifs found in SF1 helicases, but motif 3 is different and the locations of all the motifs differ: Domain 1A contains motifs 1, 1a, and 2; Domain 2A contains motifs 3, 4, 5, and 6.

Helicase motifs 1, 2 and 6 are common to helicases of all Superfamilies and contain conserved residues that are involved in the binding and hydrolysis of NTPs. Helicase motifs 1 and 2 are closely related to Walker A and B boxes associated with NTP hydrolysis (Walker *et al.* 1982). Motif 6, referred to as motif R in SF3, SF4, SF5 and SF6 helicases, contains an arginine finger that is involved in energy coupling (Scheffzek *et al.* 1997).

The *bs*PcrA helicase is well characterised and the roles of some individual amino acids of the helicase motifs have been uncovered. The crystal structures of *bs*PcrA show that Q254 of motif 3, R287 of motif 4 and R610 of motif 6 all contact the γ phosphate of AMPPNP. E571 of motif 5 interacts with the ribose of the AMPPNP and D233 and E224 of motif 2 and T38 of motif 1 contact the magnesium ion. The position of this magnesium ion is filled by K37 of motif 1 in the apo and ADP bound *bs*PcrA structures. When AMPPNP is bound K37 moves out of this position to a position where it can stabilise ATP-hydrolysis intermediates (Subramanya *et al.* 1996; Velankar *et al.* 1999). The crystal structures also show that F64 of motif 1a, Y257 and W259 of motif 3, R359 and N361 of motif 4 and H565 of motif 5 interact with DNA. Motif 3 acts as bridge between the ATP binding site and the DNA binding sites, Q254 interacts with the ATP while Y257 and W259 interact with the DNA (Subramanya *et al.* 1996; Velankar *et al.* 1999). Mutating the DNA interacting amino acids to alanine reduces the helicase activity of *bs*PcrA dramatically, but ATP hydrolysis activity is unaffected (Dillingham *et al.* 1999).

1.6 Further classification

An extension to the classification of helicases by Gorbalenya and Koonin was recently published (Singleton *et al.* 2007), which further classifies helicases by biochemical and mechanistic properties. The consequence of this is that a 'putative helicase' cannot be fully classified until it has been biochemically characterised. Where directionality of translocation can be determined, a 3' – 5' helicase is type A and a 5' – 3' helicase is type B. The SF1, SF2 and SF6 helicases contain both types A and B whereas at present all characterised SF3 helicases are type A and all SF4 and SF5 helicases are type B. Further to this, helicases that translocate along single-stranded nucleic acid are classified as type α and those that translocate along double-stranded nucleic acid as type β (see Figure 1.5).

Figure 1.5. Further classification of helicases based upon biochemical properties

A) Type Aα (brown) and Bα (grey) helicases

B) Type Aβ (brown) and Bβ (grey) helicases

Figure based upon a figure by Singleton *et al.* (Singleton *et al.* 2007).

For example, *bs*PcrA is a Superfamily 1 type Aα (SF1Aα) helicase because it translocates in a 3' – 5' direction along single-stranded DNA.

1.7 Accessory domains and proteins

The core domains of helicases do not necessarily provide the means to disrupt duplex DNA, but rather they translocate nucleic acid. Other accessory domains can provide the means to disrupt the duplex nucleic acid ahead of the translocating 'core domains'. For example the 1B and 2B domains of *bs*PcrA has been shown to be important for disrupting duplex DNA (Soultanas *et al.* 2000; Velankar *et al.* 1999) (see Chapter 3 for further details). The accessory domains do not need to be contained within the same polypeptide. In the case of *E.coli* RecBCD, two helicases, RecB and RecD, translocate along opposite strands of the duplex DNA being unwound, while RecC splits open the duplex DNA using a molecular pin and binds to the RecB and RecD subunits (Singleton *et al.* 2004).

There are also a wide variety of other accessory domains and partner proteins that are involved in targeting the activity of the helicase to a specific substrate or in providing further catalytic activities (Singleton and Wigley 2002). For example the Bacteriophage T7 gene 4 helicase-primase monomer comprises a helicase domain and a primase domain (Donmez and Patel 2006), and the helicase activity of *Bacillus stearothermophilus* DnaB protein is stimulated by the DnaG primase (Bird *et al.* 2000).

1.8 Models of helicase activity

1.8.1 Active and passive mechanisms

An active helicase is conventionally one that participates in the unwinding of duplex nucleic acid ahead of, or at, the fork of duplex DNA and the separated DNA strands (Lohman and Bjornson 1996). A passive helicase does not participate in the

unwinding of duplex nucleic acid but instead waits for thermal fraying at the fork and then traps the single-stranded nucleic acid thus preventing reannealing of the strands (Lohman and Bjornson 1996). All of the helicases characterised to date appear to actively participate in unwinding DNA (Singleton *et al.* 2007).

There are two reports contrary to this that describe the Hepatitis C virus NS3 helicase, and the Bacteriophage T7 gene product 4 (T7 gp4) helicase as passive helicases that translocate DNA in a random Brownian motion fashion (Levin *et al.* 2005; Stano *et al.* 2005). The NS3 helicase is a Superfamily 2 helicase, a non hexameric ring helicase, and it is hard to imagine how a random Brownian motion can occur with out dissociation of the helicase from the nucleic acid substrate, in which case it hard to see how the helicase would then trap single-stranded DNA produced by thermal fraying of duplex DNA. If the helicase is bound to nucleic acid it can only translocate with out dissociating from the nucleic acid, this is a highly regulated process dependant on the hydrolysis of NTPs (Lohman and Bjornson 1996). It is easy to imagine DNA slippage, i.e. a base not being translocated upon NTP hydrolysis, but it not easy to see how it would reverse direction in a random Brownian motion. The crystal structure of the helicase domain of the NS3 helicase in complex with DNA shows that there are interactions of the phosphodiester backbone of the DNA with the RecA-like fold core domains (Kim *et al.* 1998). The authors note two amino acids that act like 'book ends' on either side of the cleft between the 1A and 2A domains of the helicase. The authors suggest that the cleft between domains 1A and 2A will open and close in an NTP dependant manor, as is the case for *bs*PcrA, and will enable single-stranded nucleic acid translocation in one direction. It hard to see how this can fit with a random Brownian motion of the helicase. In the case of the T7 gp4 helicase, which is a hexameric ring it is

conceivable that the ring acts as a clamp to keep the core helicase domains in proximity to the DNA during a random Brownian motion, as the core domains would not need to stay bound to the DNA. However, a recent study that set out to determine kinetically whether the T7 gp5 helicase is an active or passive helicase concludes that enzyme must be an active helicase (Johnson *et al.* 2007). The structure of the T7 gp4 helicase also suggests a mechanism for DNA translocation that requires at least two of six the RecA-like core domains to be bound to single stranded DNA at anyone time (Singleton *et al.* 2000), making random Brownian motion of the helicase unlikely.

1.8.2 Superfamily 1 helicases

There are two main models of how SF1 helicases function: the inchworm model and the active rolling model. Both of these models were originally devised to explain data relating to the *E.coli* Rep helicase. Yarranton and Gefter proposed the inchworm model to explain the very first biochemical data gathered from studying the *E.coli* Rep helicase (also known as helicase III) (Yarranton *et al.* 1979a, b; Yarranton and Gefter 1979). The active rolling model was later proposed to explain further biochemical data relating to the *E.coli* Rep helicase (Wong and Lohman 1992).

1.8.3 The inchworm model

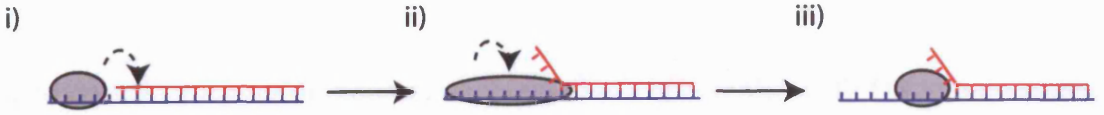

Figure 1.6. The inchworm model of helicase action

The helicase, bound to single stranded DNA, extends into, and unwinds, the duplex DNA before the rest of the helicase translocates along the DNA to catch up with the front part.

Studies of the SF1A *bs*PcrA and *E.coli* UvrD helicases, and the SF1B *E.coli* Dda helicase have provided the best evidence for the inchworm model. The inchworm model for the mechanism of the *bs*PcrA helicase activity is described in detail in Chapter 4 (see section 4.4.2.1) and shows very clearly how a monomer of *bs*PcrA could unwind duplex DNA. Biochemical evidence also points to *bs*PcrA working as a monomer. 1) no oligomeric states of *bs*PcrA have been observed (Bird *et al.* 1998). 2) monomers of *bs*PcrA translocate single stranded DNA (Dillingham *et al.* 2000, 2002). 3) amino acid mutations in the 2B domain, the domain proposed to destabilise duplex DNA ahead of the 'motor domains', reduce helicase activity (Soultanas *et al.* 2000; Velankar *et al.* 1999).

A recent study implies that *bs*PcrA is only active when part of a higher order complex because monomeric *bs*PcrA cannot unwind duplex DNA under single turnover conditions (Niedziela-Majka *et al.* 2007). However, the *Bacillus subtilis* YxaL protein and the *Staphylococcal* pC221 RepD protein both increase the processivity of *bs*PcrA *in vitro*, without modifying the DNA-dependent ATP hydrolysis or DNA binding activities of the enzyme (Noirot-Gros *et al.* 2002; Soultanas *et al.* 1999). Although these observations do not reveal anything about the

oligomeric state of the helicase, it is possible that accessory proteins could be modulating the activity of monomeric helicases.

The crystal structures of monomeric *E.coli* UvrD in complex with DNA and various nucleotide analogues were recently determined (Lee and Yang 2006). The structure of UvrD shows structural homology to PcrA and the mechanism of DNA unwinding appears to work in a very similar fashion.

The *E.coli* Dda helicase is one of the best-characterised SF1B helicases, despite the lack of structural information. Single turnover experiments show that monomers of the enzyme can unwind DNA (Nanduri *et al.* 2002) and that the Dda protein exists as a monomer and does not form oligomeric states (Morris *et al.* 2001). However, when multiple monomers are bound to the tail of a single-stranded tailed duplex DNA substrate the rate of translocation and helicase activity increases compared to a single monomer (Byrd and Raney 2004, 2005). Additionally, multiple bound Dda monomers are required to displace single-stranded binding protein from DNA (Byrd and Raney 2006). These data indicate that adjacent Dda monomers on the single-stranded DNA can either replace the active DNA-unwinding monomer upon dissociation and prevent reannealing of the separated strands, or that the monomers function cooperatively. This has led to the proposal of the cooperative inchworm model (Mackintosh and Raney 2006; Spurling *et al.* 2006) (see Figure 1.7).

Figure 1.7. The cooperative inchworm model

A) The helicase monomer unwinds a few base pairs before dissociating and the separated DNA strands reanneal. B) More than one helicase monomer binds to the DNA and when the lead monomer unwinding the DNA dissociates the other monomers continue to unwind the DNA. C) More than one helicase monomer binds to the DNA and work in a cooperative fashion allowing the lead monomer to unwind the DNA. Figure based on a figure by Mackintosh *et al.* (Mackintosh and Raney 2006).

1.8.4 The active rolling model

The *E.coli* Rep helicase is monomeric in solution (Lohman *et al.* 1989) and binds one molecule of ATP (Moore and Lohman 1994a, b). Chao *et al.* report that dimerisation of Rep is induced upon DNA binding (Chao and Lohman 1991) and Rep oligomers are required to initiate DNA unwinding (Cheng *et al.* 2001; Ha *et al.* 2002). The study of allosteric effects of nucleotide binding present five possible DNA-bound states of a Rep dimer, three of which are stabilised during the ATP hydrolysis cycle (Wong and Lohman 1993) (see Figure 1.8).

Figure 1.8. The DNA bound states of a Rep dimer

The five possible DNA bound states of a Rep dimer. P = protein, D = double-stranded DNA and S = single-stranded DNA. Stabilising co-factors are indicated, P_2D and P_2D_2 are not stable formations. Figure based on figure by Lohman *et al.* (Lohman and Bjornson 1996).

The active rolling model is put forward to explain these data (see Figure 1.9).

Figure 1.9. The active rolling model for the *E.coli* Rep helicase

Both the subunits of the Rep dimer bind to single-stranded DNA in the absence of ATP and ADP. Upon ATP binding one of the Rep monomers binds to duplex DNA, effectively translocating DNA. Upon ATP hydrolysis the duplex is destabilised and the Rep monomer binds to single-stranded DNA in the absence of ATP and ADP. Figure based on figure by Lohman *et al.* (Lohman and Bjornson 1996).

The *E.coli* Rep, *E.coli* UvrD and *bs*PcrA proteins show structural homology to one another (Korolev *et al.* 1997; Lee and Yang 2006; Subramanya *et al.* 1996). It seems unlikely that the same structural motifs work in such vastly differing models of helicase action. The active rolling model makes several assumptions about helicase activity. Firstly, the active rolling model requires SF1 helicases to work at least as dimers, however the Bacteriophage T4 Dda helicase (Nanduri *et al.* 2002), the F plasmid TraI (DNA helicase I) (Sikora *et al.* 2006) and the *E.coli* RecQ (Zhang *et al.* 2006b) helicase have been shown to function as monomers. Secondly, the active rolling model requires a step size of at least the length of a Rep monomer plus the

length of the extruded loop between the dimers. This exceeds both the largest experimental and the largest theoretical step size determined to date (see section 1.2.6). Thirdly, there is no obvious mechanism to confer directionality of unwinding, an active rolling helicase should be able to translocate bi-directionally as no single monomer is constantly bound to DNA and the Rep helicase shows clear directionality. Finally, the suggestion that Rep dimerises upon DNA binding may not be true. The crystal structure of Rep in complex with single-stranded DNA substrate, does not show any indication of Rep dimerisation because there are no interactions between the two Rep molecules bound to the DNA. Instead the length of DNA used both in the crystal structure and by Chao *et al.* is long enough to bind two Rep molecules. In order to say definitively that Rep dimerises upon DNA binding, shorter DNA needs to be used in the assays. This would mean that only one Rep molecule could bind to one molecule of DNA, so any apparent dimerisation of Rep would be real, i.e. protein interactions between Rep molecules would have to exist.

The active rolling model can explain the experimental data gained from studying the Rep helicase. However, a co-operative inchworm model explains most of the observations of all SF1 helicases and would also fit with observations of the Rep helicase. The cooperative inchworm model could also explain the fact that the 2B domain of the Rep helicases inhibits the helicase activity of a Rep monomer and the proposed mutual sequestering of the 2B domains of a Rep dimer activates the helicase (Brendza *et al.* 2005). An important point here is that the 2B deletion mutant of Rep unwinds DNA as a monomer. This shows that although Rep may prefer to work as a dimer, the core domains and the 1B domain are sufficient for monomeric helicase activity for monomeric Rep.

1.9 Superfamily 1A

The best-characterised SF1Aα proteins are the *E.coli* Rep and UvrD helicases and *bs*PcrA helicase. The crystal structures of these proteins have all been determined and show that they share the common RecA-like tandem folds (Korolev *et al.* 1997; Lee and Yang 2006; Subramanya *et al.* 1996). The mechanism for single-stranded DNA translocation has been studied in depth and the proposed model of *bs*PcrA translocation is explained in detail in Chapter 4 (see section 4.4.2.1).

The mechanism for breaking open duplex DNA is still not entirely clear. The 2B domain of *bs*PcrA is implicated in disruption of duplex DNA ahead of the motor domains, but the 2B domain of Rep is not required for helicase activity and indeed inhibits it (Brendza *et al.* 2005; Cheng *et al.* 2002).

1.10 Superfamily 1B

1.10.1 *E.coli* RecD

The *E.coli* RecD helicase is a 67KDa protein that forms a complex with the *E.coli* RecB and RecC proteins to form the RecBCD complex (Singleton *et al.* 2004). In eubacteria the RecBCD complex processes potentially lethal double-strand breaks within the genome. The DNA is digested by RecBCD using a combination of helicase and endonuclease activity. RecBCD then loads RecA onto to the DNA in order to initiate DNA repair by homologous recombination.

The *E.coli* RecD and RecB subunits of RecBCD are ssDNA-dependent ATPases (Chen *et al.* 1997; Hickson *et al.* 1985). In addition, RecB is a SF1Aα helicase (Boehmer and Emmerson 1992) and RecD is a S1Bα helicase (Dillingham *et al.*

2003). These observations fit the model of RecBCD being a 'bipolar' helicase with RecB and RecD translocating opposite strands of the duplex DNA being unwound in an inchworm helicase mechanism. This model is supported by the structure of *E.coli* RecBCD in complex with duplex DNA (Singleton *et al.* 2004)(see Figure 1.10). The RecB and RecD helicases are bound to opposite strands of the DNA, and will both translocate in opposite polarities along the opposite DNA strands and hence travel in the same overall direction.

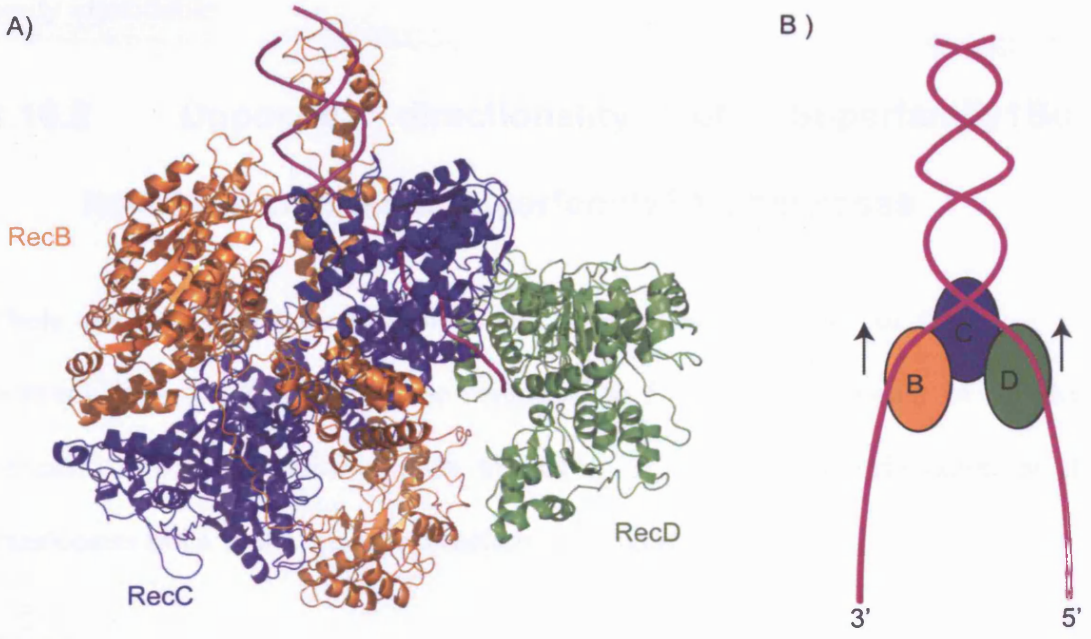

Figure 1.10. The structure of *E.coli* RecBCD complex

A) The structure of the *E.coli* RecBCD complex. B) A model of single-stranded DNA translocation by the RecB and RecD subunits; the RecC subunit splits open the duplex DNA.
The RecB subunit is coloured orange, the RecC subunit blue, and the RecD subunit green.

As described earlier there are two modes of helicase action, active and passive. The structure of the *E.coli* RecBCD complex reveals a pin in RecC that is positioned at the base of the duplex. Although the structure is by nature static, it is easy to imagine the RecB and RecD helicases translocating the opposite single-stranded DNA

strands using the pin as a mechanical device over which the duplex DNA could be split open, suggesting an active DNA unwinding mechanism.

All of the three subunits of the *E.coli* RecBCD complex contain the tandem RecA-like fold core domains (Singleton *et al.* 2004). In the case of RecC this is surprising as RecC does not contain any of the helicase motifs indicative of the RecA-like fold core domains. The 1A and 2A domains of both the RecB and RecD subunits are easily identifiable.

1.10.2 Opposite directionality of Superfamily1Bα helicases compared Superfamily1Aα helicases

There are two simple models that could explain the difference of the 3' – 5' directionality of a SF1Aα helicase compared to 5' – 3' directionality of SF1Bα helicase. Either the helicase binds the DNA in the opposite orientation or it translocates DNA in the opposite direction (see Figure 1.11).

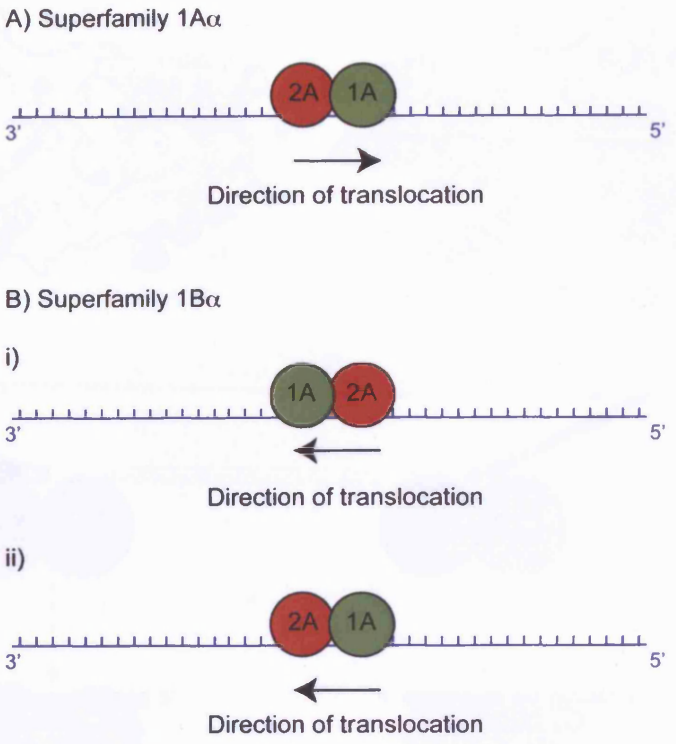

Figure 1.11. Models of how the 3' – 5' directionality of a Superfamily 1Aα helicase could be reversed to 5' – 3'

A) Translocation of a SF1A helicase showing the orientation of the 1A and 2A domains relative to the direction of translocation. B) The translocation of a SF1A helicase could be reversed in a SF1B helicase by i) binding the DNA in the opposite orientation and keeping the same translocation mechanism, or ii) by binding the DNA in the same orientation and changing the mechanism of translocation.

The structure of the *E.coli* RecBCD complex suggests that the change in directionality is not due to SF1Bα helicases binding DNA in the opposite orientation to SF1Aα helicases (Singleton *et al.* 2004). The 5' tail of the unwound single-stranded DNA binds firstly to the 1A domain of the RecD1 helicase then to the 2A domain, which is analogous to the binding of the 3' tail of single-stranded DNA to the 2A domain of the SF1Aα helicase *bs*PcrA and then to the 1A domain. In both cases the polarity of the bound DNA is the same across the 1A and 2A domains of *E.coli* RecD and *bs*PcrA (see Figure 1.12).

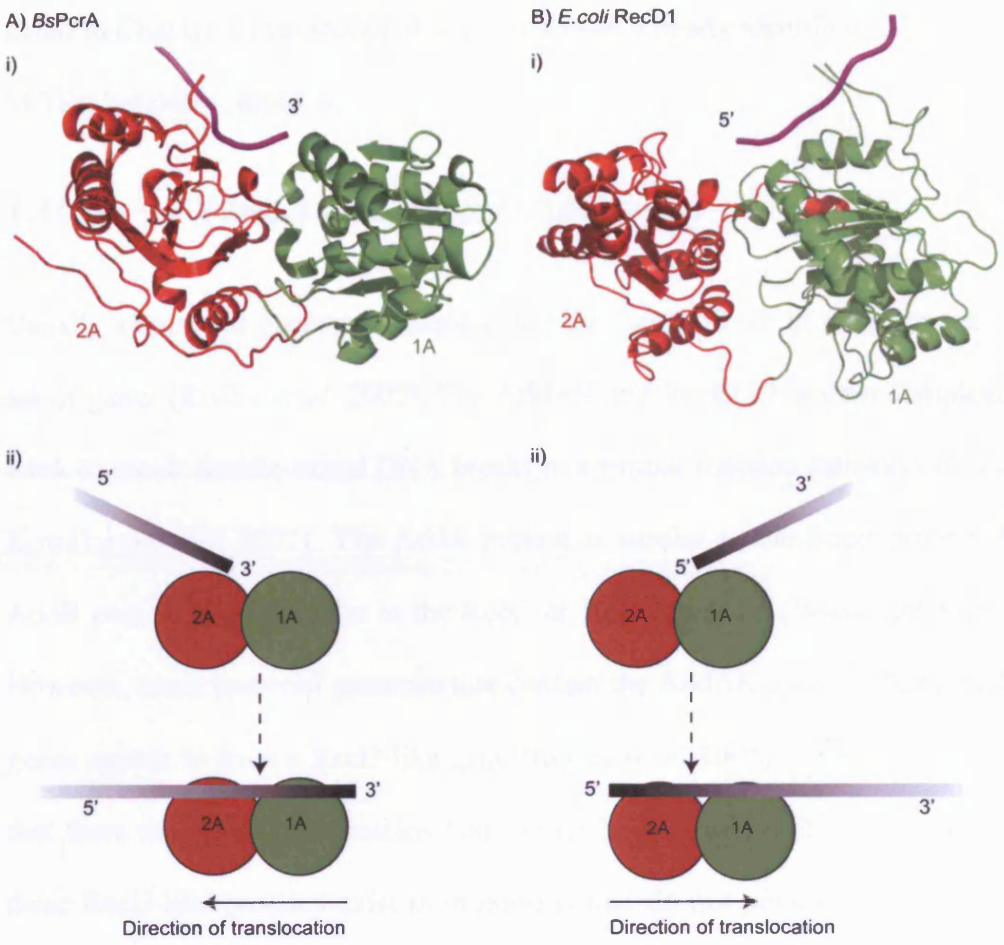

Figure 1.12. The orientation of DNA binding by Superfamily 1Aα and Superfamily 1Bα helicases

A) i) The structure and, ii) a schematic, of the binding of single stranded DNA to the 1A and 2A domains of *bs*PcrA. B) i) The structure and ii) a schematic of the binding of single-stranded DNA to the 1A and 2A domains of *E.coli* RecD. The 1A and 2A domains of the resulting bound DNA states have the same orientation relative to the asymmetric single-stranded DNA.

The shared orientation of the single-stranded DNA binding to the motor domains by *E.coli* RecD1 and *bs*PcrA is highly significant as it suggests the mechanism of RecD translocation is reversed and that the molecular details of the helicase mechanism of SF1Aα and SF1Bα helicases are not the same.

Sequence comparisons of helicase motifs 1a and 3 between SF1Aα and SF1Bα helicases reveal a clue into how the helicase mechanisms differ. This is discussed in

detail in Chapter 4 (see section 4.4.2) along with a newly identified motif common to SF1Bα helicases, motif A.

1.10.3 Type 1 and 2 RecD helicases

Usually a bacterial genome contains either the RecBCD set of genes or the AddAB set of genes (Rocha *et al.* 2005). The AddAB and RecBCD protein complexes both work to repair double-strand DNA breaks in a similar reaction pathways (Chedin and Kowalczykowski 2002). The AddA protein is similar to the RecB protein, but the AddB protein is not similar to the RecC or RecD proteins (Wang and Julin 2004). However, some bacterial genomes that contain the AddAB genes and not the RecBC genes appear to have a RecD-like gene (Rocha *et al.* 2005). It has become apparent that there are RecD-like proteins that cannot be part of RecBCD complex because these RecD-like proteins exist in organisms that do not possess the RecB or RecC genes. The RecD enzymes that are likely to exist as part of the RecBCD complex are referred to as type 1 RecD (RecD1) helicases and the RecD-like enzymes that are not part of the RecBCD complex are referred to as type 2 RecD (RecD2) helicases (Rocha *et al.* 2005). Table 1.1 summarises the types of bacteria that contain the AddAB or RecBCD genes and the RecD-like (RecD2) genes (Rocha *et al.* 2005).

Table 1.1. Summary of the presence of the Add AB, RecBCD1 and RecD2 genes in bacterial genomes

Type of bacteria	Add AB	RecBCD1	RecD2
γ-proteobacteria	-	+	-
β- proteobacteria *α-proteobacteria*	+	-	-
δ- proteobacteria, Cyanobacteria, Acinobacteria	-	-/+	-/+
ε- proteobacteria	-	-	-
Firmicutes	+	-	+
Spirochetes	-	-/+	-
Chlamydiae	-	+	+

The *Chlamydiae* species and *Deinococcus radiodurans* are significant when looking at the presence or absence of the AddAB, RecBCD1 and RecD2 genes. The *Chlamydiae* are unusual in the fact that they contain the RecBCD1 genes as well as a RecD2 gene. *Deinococcus radiodurans* is unusual because it does not contain either the Add AB or RecBCD1 set of genes and most bacteria appear to have either one set or the other (Rocha *et al.* 2005). Rocha *et al.* have carried out a phylogenetic analysis of the RecD1 and RecD2 proteins (see Figure 1.13). It shows that the RecD1 family is not only distinguished from the RecD2 family by the presence or absence of the RecBC genes within the genome of an organism, but also by the sequence of the RecD gene, furthering the argument that there are two distinct types of RecD proteins.

Figure 1.13. An un-rooted phylogenetic tree of the RecD1 and RecD2 proteins

The dotted line splits the genomes containing RecBCD1 from those containing only RecD2. *C.acetobutylicum, Clostridium acetobutylicum; C. tepidum, Chlorobium tepidum; C.violaceum, Chromobacterium violaceum; D. vulgaris, Desulfovibrio vulgaris; E. carotovora, Erwinia carotovora; E. faecalis, Enterococcus faecalis; G.sulfurreducens Geobacter, sulfurreducnes; L.plantarum, Lactobacillus plantarum; L. lactis, Lactococcus lactis; P.multocida, Pasteurella multocida; M. mobile, Mycoplasma mobile; M. florum, Mesoplasma florum; M. mycoides, Mycoplasma mycoides; M.pulmonis, Mycoplasma pulmonis; N.meningitidis, Neisseria meningitides; P. maritima, Prochlorococcus maritima; S. enterica, Salmonella enterica; S.oneidensis, Shewanella oneidensis; X.fastidiosa, Xylella fastidiosa.* Figure based upon a figure by Rocha *et al.* (Rocha *et al.* 2005).

The *E.coli* RecD1 helicase does not interact with the duplex portion of the DNA being unwound by the RecBCD complex (Singleton *et al.* 2004). The N-terminal domain of the *E.coli* RecD1 helicase interacts with RecC. This domain is not conserved in RecD2 helicases, probably because RecD2 helicases do not interact with a RecC protein. The role of the N-terminal region of RecD2 helicases, which tend to be larger than the N-terminal domain of RecD1 helicases, is yet to be fully understood. It may be involved in interacting with the duplex region of DNA.

1.10.4 *Deinococcus radiodurans* RecD2

As mentioned earlier, *Deinococcus radiodurans* is unusual because its genome does not contain either the AddAB or RecBCD1 genes, although it does contain a RecD2 gene that encodes for a 77 KDa protein (White *et al.* 1999). The AddAB and RecBCD1 pathways are involved in DNA repair and so the absence of the these genes is even more odd when coupled with the fact that *Deinococcus radiodurans* is one of the most DNA damage resistant organisms characterised to date (Battista 1997; Battista *et al.* 1999; Minton 1994). The *Deinococcus radiodurans* genome does contain a RecA gene (Gutman *et al.* 1994; Kim *et al.* 2002) despite not having the AddAB or RecBCD1 complexes to target loading RecA onto areas of DNA damage.

The precise *in vivo* role of the type 2 RecD proteins is not known. The *Deinococcus radiodurans* RecD2 helicase is the only type 2 RecD enzyme to have had its *in vivo* function investigated and has recently been shown to play a role in DNA repair and/or homologous recombination (Servinsky and Julin 2007), as well as being reported to be involved in an antioxidant pathway (Zhou *et al.* 2007).

1.10.5 Eukaryotic Superfamily 1B helicases

Several eukaryotic SF1B helicases have been studied are associated human diseases.

1.10.5.1 The Pif1-like helicases

Four members of the Pif1-like helicases have been studied in detail: two budding yeast helicases, *sc*Pif1 and *sc*Rrm3p and the Pif1p helicase from the fusion yeast *Schizosaccharomyces pombe* (*sp*) and humans. Like the RecD helicases, the Pif1-like proteins all display single-stranded DNA dependent ATP hydrolysis and 5' – 3'

helicase activities (Ivessa *et al.* 2002; Lahaye *et al.* 1993; Zhang *et al.* 2006a; Zhou *et al.* 2002). The fission yeast *sp*Pif1p helicase gene is essential, as is the helicase function of the protein (Zhou *et al.* 2002). The *sp*Pif1p helicase is implicated in telomere maintenance, but is also essential in cells with circular chromosomes, so this cannot be the essential function, and the helicase function of *sp*Phf1p is implicated in Okazaki fragment maturation (Boule and Zakian 2006). The role of the human Pif1p helicase is still not clear, but it seems to be involved in telomere regulation (Boule and Zakian 2006; Zhang *et al.* 2006a).

1.10.5.2 The human Upf1 helicase

The human Upf1 (hUpf1) helicase is involved in nonsense-mediated mRNA decay (NMD), a process that recognises and destroys aberrant mRNAs to prevent the synthesis of truncated proteins within the cell (Baker and Parker 2004; Conti and Izaurralde 2005; Lejeune and Maquat 2005). The human and budding yeast Upf1 helicases both possess RNA binding, RNA-dependent ATP hydrolysis and 5' – 3' helicase activities, these activities are required for NMD (Bhattacharya *et al.* 2000; Czaplinski *et al.* 1995; Weng *et al.* 1996a, b).

The structure of the hUpf1 helicase was recently solved (Cheng *et al.* 2007) and is the first example of a human SF1 helicase structure. As expected the hUpf1 structure shows structural homology to the prokaryotic SF1 helicases such as *bs*PcrA, *E.coli* Rep, UvrD and RecD, and contains the RecA-like fold core domains predicted by the presence of the seven Superfamily 1 helicase motifs (Cheng *et al.* 2007). This is a clear indication that the SF1 helicase structure and, therefore, the associated helicase mechanisms are conserved across all domains of life.

The inhibition of NMD by inhibiting hUpf1 protein expression has been used to reverse a phenotype, the loss of the ability to create functional extra cellular matrix, of Ullrich diseased fibroblasts taken from a patient (Usuki *et al.* 2004; Usuki *et al.* 2006). This phenotype of the fibroblasts contributes toward to a clinical symptom of the patient: a form muscular dystrophy (Usuki *et al.* 2006).

1.10.5.3 The BRCA1-associated C-terminal helicase 1 (BACH1)

The BRCA1-associated C-terminal helicase 1 (BACH1) helicase (also known as FancJ and Brip1) interacts with the BRCT domain of the Breast Cancer Associated Protein 1 (BRCA1)(Cantor *et al.* 2001). Defects in the BRCA1 gene are linked to the development of hereditary breast, ovarian, and other cancers (Alberg *et al.* 1999; Futreal *et al.* 1994; Miki *et al.* 1994; Nathanson *et al.* 2001). When the helicase function of BACH1 is knocked out, the double-strand break repair activity mediated by the BRCA1/BACH1 complex is impaired (Cantor *et al.* 2001; Kumaraswamy and Shiekhattar 2007). Further to this two germ-line mutations in the helicase domain of BACH1 have been observed in patients with early onset breast cancer (Cantor *et al.* 2001).

1.11 Thesis outline

This thesis is in two parts. The first part of the thesis (Chapter 3) describes attempts to utilizes the protein splicing properties of inteins to create a full-length *bs*PcrA protein from two halves *in vitro*. This was to be the first step in creating a *bs*PcrA deletion mutant missing the 2B domain that could then be used to investigate the disputed role of the 2B domain of SF1 helicases.

The second part of this thesis (Chapters 4 and 5) aims to further the understanding of the helicase mechanism of SF1B helicases. Chapter 4 describes the identification of a type 2 RecD helicase from *Deinococcus radiodurans* as a suitable target for study and goes on to described the purification of *Deinococcus radiodurans* RecD2 (*dr*RecD2). Chapter 4 also characterises the biochemical properties of the *dr*RecD2 helicase and describes the study of mutant *dr*RecD2 helicases to examine the SF1B helicase mechanism. Chapter 5 describes the crystallisation of *dr*RecD2 in complex with a 5'-tailed duplex DNA substrate.

2 Materials and methods

2.1 Standard materials and methods

2.1.1 Standard methods

Molecular biology techniques were carried out as described in standard laboratory manuals (Ausubel *et al.* 1988; Sambrook *et al.* 1989). Where enzymes were used for the manipulation of DNA the supplier's protocol was followed. Any deviation from the standard protocols is noted. Concentrations of protein and DNA were estimated using the theoretical extinction coefficients.

2.1.2 Standard materials

Unless otherwise stated all chemicals were obtained from VWR International, Fisher Scientific and Sigma-Aldrich. Adenosine [γ-^{32}P] triphosphate was obtained from GE Healthcare. Chromatography columns and media were obtained from GE Healthcare and any exceptions are noted. Oligonucleotides were synthesised by Sigma-Genosys and an in-house service.

2.2 Microbiological methods

2.2.1 *E.coli* strains

The *E.coli* strains used during the work described in this thesis are listed in Table 2.1.

Table 2.1. Summary of *E.coli* expression strains

Strain	Features
E.coli XL1-blue (Bullock *et al.* 1987) (Stratagene)	Used for cloning. Lacks RecA.
E.coli B834 (DE3) (Studier and Moffatt 1986) (Novagen)	Used for protein and seleno-methionine protein expression. Lacks the ability to synthesise methionine. Deficient in both lon and ompT proteases.
E.coli BL21 (DE3) (Novagen)	Used for protein expression. Derived from B843 (DE3) cells.
E.coli TUNER (Novagen)	lacZY deletion mutants of BL21. The lac permease (lacY) mutation allows uniform entry of IPTG into all cells of the culture. By adjusting the concentration of IPTG, protein expression can be regulated.

Rare codon plasmids (helper expression plasmids) are listed in Table 2.2.

Table 2.2. Summary of helper expression plasmids

Rare codon plasmid	Features
Codon + (pRIL, Stratagene)	Supplies tRNAs for the codons AGA, AGG, AUA and CUA. Chloramphenicol resistant.
Codon 2 + (Provided by Steve Sandler)	Supplies tRNAs for the codons AGG and AGU. Spectinomycin resistant.
Codon 3 + (pRP, Stratagene)	Supplies tRNAs for the codons AGG, AGA and CCC. Chloramphenicol resistant.
Rosetta (pRARE, Novagen) (Novy *et al.* 2001)	Supplies tRNAs for the codons AGG, AGA, AUA, CUA, CCC, and GGA. Chloramphenicol resistant.

When expression of a protein is referred to both the bacterial strain and the helper plasmid are both quoted. For example, a protein expressed in the *E.coli* expression

strain BL21(DE3) with the helper plasmid codon 2+ is quoted as being expressed in BL21 2+.

2.3 DNA methods

2.3.1 Oligonucleotide purification

The oligonucleotide was suspended in denaturing buffer A (50 mM phosphate buffer pH 6.8, 10 mM sodium hydroxide) with 100 mM sodium chloride and was loaded onto a Source 15Q column equilibrated in denaturing buffer. After washing away unbound material the oligonucleotide was eluted with a 30 column-volume gradient from 0.1 M to 2.0 M sodium chloride in denaturing buffer A. The oligonucleotide was exchanged into water using a NAP25 drip column, concentrated using a centricon (Amicon) with an appropriate molecular weight cut off and stored at -20 °C.

2.3.2 Annealing oligonucleotides

Equimolar amounts of complementary oligonucleotide in water were mixed in annealing buffer (50 mM phosphate buffer pH 6.8, 100 mM potassium chloride, 10 mM magnesium chloride) before being heated to 95 °C in a programmable thermal cycler. The thermal cycler was programmed to cool the mixture from 95 °C to 4 °C over 18 hours. The resulting annealed oligonucleotides were stored at -20 °C.

2.3.3 Cloning

The target gene to be cloned was used as the template for a polymerase chain reaction (PCR). The primers were designed so that the resulting PCR product would contain non-complementary flanking sequences. The flanking sequences contained

restriction sites which were also contained within an expression vector plasmid. The remainder of the primer (i.e. the region complementary to the template sequence) was designed to have a theoretical melting point of 65 °C . Polymerase chain reactions were carried out using the Expand high fidelity PCR system (Roche) according to the manufacturers protocol (an exception is noted later) and the product was purified using the QIAquick PCR purification kit (Qiagen). The purified product and the target expression plasmid were digested with the appropriate restriction enzymes and were separated by agarose gel electrophoresis and purified using the QIAquick gel extraction kit (Qiagen). The restriction digest results in complementary overhangs between the PCR product and the expression plasmid. The digested plasmid was mixed with an excess of the product (insert) in a ligation reaction with DNA ligase (New England Biolabs) before being transformed into supercompetent *E.coli* XL1-Blue cells (Stratagene) via electroporation. The *E.coli* were plated on to agar with the appropriate antibiotics and incubated at 37 °C overnight. Up to six colonies were picked and cultured in 5 ml of LB with appropriate antibiotic cultures and incubated at 37 °C with shaking overnight. The plasmids were purified using the QAI plasmid mini preparation kit (Qiagen). The presence of the correct insert in the plasmid was confirmed by restriction digest with the appropriate restriction enzymes analysed by agarose-gel electrophoresis and by sequencing (in-house service). An exception to this was the PCR buffer for PCR of the wild type *dr*RecD2 gene using the *Deinococcus radiodurans* genomic DNA as the template. The PCR also contained 2.5 % w/v DMSO; the DNA denaturing steps were 1 minute at 99 °C instead of 30 seconds at 95 °C and the initial denaturing step was 10 minutes at 99 °C instead of 5 minutes at 95 °C.

All the constructs used in this thesis are listed in Table 2.3.

Table 2.3. Constructs of this thesis

Gene cloned	Expression plasmid	Restriction sites used	Notes
Wild-type *Deinococcus radiodurans* RecD2	pET 22b	*Nde* 1 *Xho* 1	C-terminal hexa-histidine tag
Deinococcus radiodurans $RecD2_{new}$ and RecD2 mutants	pET 22b	*Nde* 1 *Xho* 1	C-terminal hexa-histidine tag
E.coli RecD1	pET 15b	*Nde* 1 *Bam* H1	N-terminal hexa-histidine tag
Streptococcus pneumoniae RecD 2	pET 22b	*Nde* 1 *Bam* H1	
Bacillus cereus RecD2	pET 22b	*Nde* 1 *Bam* H1	
Lactobacillus plantarum RecD2	pET 22b	*Nde* 1 *Bam* H1	
Desulfovibrio vulgaris RecD2	pET 22b	*Nde* 1 *Bam* H1	
Methanococcus jannaschii YF19_METAJ	pET 22b	*Nde* 1 *Bam* H1	
Haemophilus Influenzae RecD1	pET 15b	*Nco* 1 *Bam* H1	
Mycobacterium smegmatis RecD1	pET 22b	*Nde* 1 *Bam* H1	
Bacillus stearothermophilus PcrA1	pTWIN1	*Nde* 1 *Sap* 1	*bs*PcrA1 was fused to the N-terminus of *Mxe*GyrAi
Bacillus stearothermophilus PcrA1	pTWIN2	*Nde* 1 *Sap* 1	*bs*PcrA1 was fused to the N-terminus of *Mth*RIR1i
Bacillus stearothermophilus PcrA (amino acids 1-302, 1-312, 1-327 and 1-359) and any associated mutants.	pTWIN1	*Nde* 1 *Sap* 1	*bs*PcrA was fused to the N-terminus of *Mxe*GyrAi
Bacillus stearothermophilus PcrA2	pTWIN1	*Sap* 1 *Pst* 1	*bs*PcrA2 was fused to the C-terminus of *Ssp*DnaBi
Archaeoglobus fulgidus PCNA	pTWIN1	*Nde* 1 *Sap* 1	*af*PCNA was fused to the N-terminus of *Mxe*GyrAi
Archaeoglobus fulgidus PCNA	pTWIN2	*Nde* 1 *Sap* 1	*af*PCNA was fused to the N-terminus of *Mth*RIR1i
Archaeoglobus fulgidus PCNA	pTWIN1	*Sap* 1 *Bam* H1	*af*PCNA was fused to the C-terminus of *Ssp*DnaBi via the *pf*PCNA linker

All the vectors used contained an ampicillin resistance gene as a selective marker. The pET vectors were obtained from Novagen and the pTWIN vectors were obtained from New England Biolabs.

2.3.4 Site-directed mutagenesis

To produce point mutants of a protein the Quick-change kit 2.0 was used according to the supplier's instructions (Qiagen).

The *dr*RecD2$^{\Delta loop}$ was created by replacing 412-419 (inclusive) of *dr*RecD2 with a glycine, using a three step PCR construction of the gene as described in Figure 2.1.

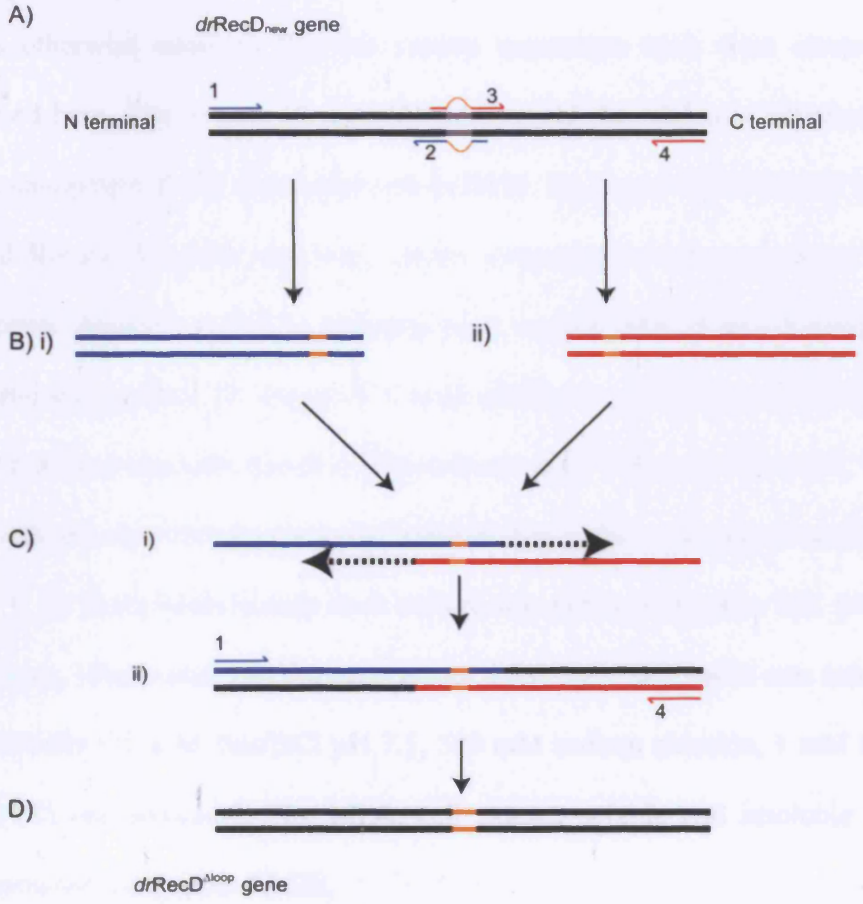

Figure 2.1. The cloning of *dr*RecD2$^{\Delta loop}$

A) A schematic of the *dr*RecD2$_{new}$ gene with the region that encodes amino acids 412-419 coloured grey. The primers to be used are also shown. The parts of the primers that are not complementary to the gene sequence and instead encode a glycine residue are coloured orange.

B) i) The resulting PCR product when primers 1 and 2 are used. ii) The resulting PCR product when primers 3 and 4 are used.

C) The PCR products from B) are mixed and i) one round of a PCR is performed in the absence of primers before ii) the addition of primers 1 and 4 to the PCR.

D) The resulting PCR product from C). The gene no longer encodes amino acids 412-419, but encodes a glycine instead.

Primers 1 and 4 were designed so that the PCR product, the *dr*RecD2$^{\Delta loop}$ gene, would contain two flanking restriction sites for cloning into an expression plasmid.

2.4 Protein methods

2.4.1 Protein expression

2.4.1.1 Protein expression trials

Unless otherwise stated in the text protein expression trials were carried out as described here. The expression plasmid undergoing the trial was transformed into electro-competent *E.coli* expression strains B834, BL21 and TUNER with the +, 2+, 3+ and Rosetta plasmids and were grown overnight on LB agar plates with the appropriate antibiotic at 37 °C. Colonies were washed into 20 ml LB cultures with antibiotic and grown at 37, 30 or 18 °C with shaking to an $O.D._{600}$ of 0.5-0.7 AU and then the culture was split into four. The cultures were induced with 1, 0.2, 0.04 mM IPTG and one control culture was not induced. The cultures were incubated at 37, 30 or 18 °C for three hours before each culture was pelleted (Hyspin 16K (Anachem) 14000 rpm, 10 minutes) and the supernatant discarded. Each pellet was resuspended in trial buffer (20 mM Tris/HCl pH 7.5, 500 mM sodium chloride, 1 mM EDTA, 1 mM DTT) and sonicated. The whole cell extract, soluble and insoluble fractions were analysed using SDS-PAGE.

2.4.1.2 Protein expression details

Protein was routinely expressed in the *E.coli* expression strains described in section 2.2.1. Expression plasmids were transformed into electro-competent *E.coli* and grown overnight on LB plates at 37 °C. Colonies were washed into 50 ml LB with antibiotic starter cultures and grown at overnight at 37 °C with shaking. 5 ml of starter culture was used to inoculate 1 L or 200 ml cultures that were incubated with

shaking at 37, 30, or 18 °C until the $O.D_{600}$ reached 0.5 -0.7 AU. The culture was then induced with 1 or 0.2 mM IPTG and incubated for a further 3, 6 or 18 hours. The cells were harvested by centrifugation (J-6B (Beckman), 4000 rpm, 15 minutes) and the cell pellets were stored at -80 °C.

Table 2.4 below lists the protein expression parameters

Table 2.4. Protein expression parameters

Protein expressed	***E.coli* strain**	**Temperature (°C)**	**[IPTG] (mM)**	**Length of induction (hrs)**	**Size of culture (Litres)**
*dr*RecD2 (from wild type *dr*RecD2 gene)	BL21 Rosetta	37	1	3	1
seleno-methionine *dr*RecD2* (from wild type *dr*RecD2 gene)	B834 Rosetta	37	1	6	1
*dr*RecD2 & *dr*RecD2 mutants (from $drRecD2_{new}$ gene)	BL21	37	1	3	1
E.coli RecD1 (Dillingham *et al.* 2003)	BL21	37	1	3	1
*bs*PcrA1-*Mxe*GyrAi	TUNER 2+	18	0.2	18	1
*Ssp*DnaBi-*bs*PcrA2	BL21	37	1	3	1
*af*PCNA-*Mxe*GyrAi & *af*PCNA-*Mth*RIR1i	B834 2+	37	1	3	0.2

* Grown in seleno-methionine media

2.4.2 Protein purification

The *E.coli* RecBC complex and the *E.coli* RecBCD complex were purified by Nicola Cook and Robert Court.

2.4.2.1 *Deinococcus radiodurans* RecD2 small scale purification

The cell pellet from a 1 L expression culture was resuspended in 20 ml buffer D1 and sonicated. The soluble fraction was clarified by centrifugation (SS-34 rotor (Sorvall), 20 minutes, 18000 rpm) and was incubated for 1 hour at 4 °C with gentle agitation with nickel-agarose beads (Qiagen) equilibrated in buffer D3 with 40 mM imizadole. The nickel-agarose beads were pelleted (Hyspin 16K (Anachem), 1000 rpm 2 minutes) and the supernatant discarded. The nickel-agarose beads were washed by being resuspended in 20 ml buffer D3 with 40 mM imizadole, then pelleted by centrifugation (Hyspin 16K (Anachem), 1000 rpm 2 minutes) and the supernatant discarded. Any *dr*RecD2 bound to the nickel-agarose beads was eluted by resuspending the beads in buffer D3 with 400 mM imizadole and then pelleting the beads by centrifugation (Hyspin 16K (Anachem), 1000 rpm 2 minutes). All the steps of the purification were analysed by SDS-PAGE.

2.4.2.2 *Deinococcus radiodurans* RecD2 and RecD2 mutants large scale purification

The cell pellet from ten 1 L cultures was resuspended in 200 ml buffer D1 and the cells were lysed under pressure using a homogeniser (Stansted). The soluble fraction was clarified by centrifugation (SS-34 rotor (Sorvall), 20 minutes, 18000 rpm). Ammonium sulphate was added to 65 % saturation (0.43 g/ml) and the mixture incubated for 20 minutes at 4 °C with stirring. The ammonium sulphate suspension was pelleted by centrifugation (SS-34 rotor (Sorvall), 20 minutes, 18000 rpm) and

the supernatant discarded. The pellet was resuspended in 200 ml buffer D2 and was loaded onto a nickel-chelating column (HiTrap chelating HP 5 ml) equilibrated in buffer D3 with 40 mM imizadole. After washing off unbound material *dr*RecD2 was eluted with a step gradient from 40 mM to 400 mM imizadole in buffer D3. Eluted fractions from the column were analysed by SDS-PAGE and those containing *dr*RecD2 were pooled. The pooled fraction was diluted in buffer D4 until the conductivity was between 5 and 10 mS cm^{-1} and was loaded onto a heparin column (HiTrap heparin HP 5 ml) equilibrated in buffer D4 with 50 mM sodium chloride. After washing away unbound material *dr*RecD2 was eluted with a 10 column volume linear gradient from 0.05 – 1.5 M sodium chloride in buffer D4. The fractions from the column were analysed by SDS-PAGE and those containing *dr*RecD2 were pooled. The pooled fraction was diluted in buffer D4 until the conductivity was between 5 and 10 S cm^{-1} and was loaded onto a Mono Q column (HR 5/5) equilibrated in buffer D4 with 50 mM sodium chloride. After washing away unbound material *dr*RecD2 was eluted with a 10 column volume linear gradient from 50 – 500 mM sodium chloride in buffer D4. The fractions from the column were analysed by SDS-PAGE and those containing pure *dr*RecD2 were pooled. The pooled fraction was dialysed against 4 L of buffer D5 over night at 4 °C. Aliquots of the *dr*RecD2 were frozen in liquid nitrogen and stored at -80 °C.

2.4.2.3 Seleno-methionine *Deinococcus radiodurans* RecD2

Seleno-methionine *dr*RecD2 was purified as described for wild-type *dr*RecD2, but with 10 mM BME added to the buffers. A fast flow nickel-chelating column (FF chelating column 5 ml) was used instead of the HiTrap chelating column.

2.4.2.4 *E.coli* RecD

E.coli RecD was purified as described by Dillingham *et al.* (Dillingham *et al.* 2003).

2.4.2.5 Small scale purification of proteins fused to either *Mxe*GryAi or *Mth*RIR1i

The pellet from a 200 ml expression culture was resuspended in 2 ml buffer I2 and sonicated. The soluble fraction was clarified by centrifugation (SS-34 rotor (Sorvall), 20 minutes, 14000 rpm) and was incubated for 1 hour at room temperature with 50 µl chitin bead slurry (New England Biolabs) equilibrated in buffer I1. The chitin beads were pelleted (Hyspin 16K (Anachem), 1000 rpm 2 minutes) and the supernatant discarded. The beads were washed in buffer I1. Any intein fusion protein bound to the beads was induced to cleave off the target protein by suspending the beads in 100 µl buffer I2 for 1 hour at room temperature. The beads were pelleted by centrifugation (Hyspin 16K (Anachem), 1000 rpm 2 minutes) and the supernatant and all the other steps of the purification were analysed by SDS-PAGE.

2.4.2.6 Large scale purification of proteins fused to either *Mxe*GryAi or *Mth*RIR1i

These purifications were carried out as described in the IMPACT-TWIN system manual (New England Biolabs). The chitin column comprised a 10 ml C 10/10 column containing chitin beads (New England Biolabs).

The cell pellets from three 1 L (or fifteen 200 ml) cultures were resuspended in 50 ml buffer I2 and lysed under pressure using a homogeniser (Stansted). The soluble fraction was clarified by centrifugation (SS-34 rotor (Sorvall), 20 minutes, 18000 rpm) and loaded onto the chitin column equilibrated in buffer I1 at room temperature. The chitin column was washed with 5 column volumes of buffer I1 before being equilibrated with 5 column-volumes of buffer I2 and incubated overnight at room temperature to induce intein cleavage. The target protein cleaved

from the intein was eluted with buffer I2 and the fractions analysed by SDS-PAGE. The chitin column was washed with buffer I4 to remove the intein bound to the chitin. The fractions containing the target protein were pooled and either aliquots were frozen in liquid nitrogen and stored at -80 °C or, if the protein was to be used for intein-mediated protein ligation, was immediately mixed with the ligation partner.

2.4.2.7 Large scale purification of proteins fused to *Ssp*DNABi

Purification was carried out as described in the IMPACT-TWIN system manual (New England Biolabs). The chitin column comprised a 10 ml C 10/10 column containing chitin beads (New England Biolabs). Buffer I3 and the chitin column were pre-chilled to 4 °C.

The cell pellet from three 1 L cultures was resuspended in 50 ml buffer I3 and lysed under pressure using a homogeniser (Stansted). The soluble fraction was clarified by centrifugation (SS-34 rotor (Sorvall), 20 minutes, 18000 rpm) and loaded onto the chitin column equilibrated in buffer I3 at 4 °C. The chitin column was washed with 5 column volumes of buffer I3 before being equilibrated with 5 column volumes of buffer I1 and incubated overnight at room temperature in order to induce intein cleavage. The target protein cleaved off the intein was eluted with buffer I1 and the fractions analysed by SDS-PAGE. The chitin column was washed with buffer I4 to remove the intein bound to the chitin. The fractions containing the target protein were pooled and aliquots were frozen in liquid nitrogen and stored at -80 °C.

2.4.3 Ligation of intein prepared polypeptides

Ligation of intein prepared polypeptides was carried out as described in the IMPACT-TWIN manual (New England Biolabs). The equimolar amounts of protein were mixed and were concentrated to 80 μM in the ligation buffer (20 mM Tris/HCl pH 7.0, 500 mM sodium chloride, 25 mM MESNA, 1 mM EDTA). The concentrated protein mix was incubated overnight at 4 °C or at room temperature and analysed by SDS-PAGE.

2.4.4 Protein/protein complex analysis by gel-filtration chromatography

Purified proteins were mixed with a molar excess of the smaller moiety, concentrated to less than 3 ml and purified on a Superdex S200 or S75 gel filtration column equilibrated in buffer GF (20 mM Tris/HCl pH 7.5, 100 mM sodium chloride, 1 mM EDTA, 1 mM DTT). Fractions were analysed by SDS-PAGE.

2.4.5 Dynamic light scattering

Dynamic light scattering was performed using a Zeta-sizer Nano series (Malvern) instrument according to the manufacturer's instructions. 1 ml of protein at a concentration of 1 mg/ml in DLS buffer (20 mM Tris/HCl pH 7.5, 100 mM sodium chloride, 1 mM EDTA, 1 mM DTT) was used. The results were analysed using the software supplied.

2.4.6 Identification of proteins by mass spectrometry

Proteins were purified by SDS-PAGE and the band to be identified excised and suspended in 20 µl of water. Proteins were then identified by peptide mass mapping at the Aberdeen Proteome Facility.

2.5 Biochemical assay methods

Each experiment was repeated at least three times and the mean value taken as the result. The error was taken as the standard deviation of the mean, and where appropriate is plotted as error bars on graphs.

2.5.1 DNA substrates

2.5.1.1.1 The 5'-tailed duplex substrate

The 5'-tailed duplex comprised oligonucleotide 1 (O1) and oligonucleotide 2 (O2) (see Figure 2.2). Unless otherwise stated this is the substrate referred to as the 5'-tailed duplex substrate. The duplex part of the substrate is 20 bp long and the 5' tail is 12 nt long.

```
O1              3' CAGCTACACGTATGATGCCG 5'
O2 5' TACAGCTACCTAGTCGATGTGCATACTACGGC 3'
```

Figure 2.2. The 5'-tailed duplex substrate

If the text specifies a tail length, then O2 was modified to make the 5' tail either 10 or 8 nucleotides long.

The 40bp 5'-tailed duplex substrate comprises oligonucleotide 3 (O3) and oligonucleotide 4 (O4) (see Figure 2.3).

```
O3 3'              CAGCTACACGTATGATGCCGAGTCGCTAAGCTTACGGTAC 5'
O4 5'TACAGCTACCTAGTCGATGTGCATACTACGGCTCAGCGATTCGAATGCCATG 3'
```

Figure 2.3. 40bp 5'-tailed duplex substrate

2.5.1.1.2 The ssDNA substrate

The ssDNA substrate comprised a 15nt poly d-thymine substrate (pdT15).

2.5.1.1.3 The dsDNA substrate

The dsDNA substrate comprised oligonucleotide 1 (O1) and oligonucleotide 5 (O5) (see Figure 2.4). Unless otherwise specified in the text this is the substrate referred to as the dsDNA substrate.

```
O1 3' CAGCTACACGTATGATGCCG 5'
O5 5' GTCGATGTGCATACTACGGC
```

Figure 2.4. The dsDNA substrate

2.5.2 Radio labelling of DNA

DNA was radio-labelled by the incorporation of adenosine [γ-^{32}P] triphosphate using T4 polynucleotide kinase (New England Biolabs). Labelled oligonucleotides were purified using a Micro Biospin P6 or P30 spin column (Biorad). The concentration of radio-labelled oligonucleotide was determined, based on the counts per minute of radioactivity compared to a calibration curve of known concentrations of adenosine [γ-^{32}P] triphosphate. Radio-labelled oligonucleotide was mixed with the same cold oligonucleotide in a 1:20 ratio to produce a stock of 5 % labelled oligonucleotide. If the DNA substrate comprised two oligonucleotides, one oligonucleotide was radio-labelled. The 5 % radio-labelled oligonucleotide was annealed with the appropriate complementary oligonucleotide in equimolar amounts.

2.5.3 ATP-hydrolysis assay

2.5.3.1 Experimental setup

Coupled ATP hydrolysis assays were carried out as described previously (Pullman *et al.* 1960; Tamura and Gellert 1990). The ATPase being assayed converts ATP to ADP and then pyruvate kinase transfers a phosphate group from phosphoenolpyruvate to the ADP thus regenerating the ATP making this assay continuous with constant ATP concentration. The resulting pyruvate is converted to lactate via the oxidation of NADH to NAD^+ by lactate dehydrogenase. NADH absorbs light of wavelength 340 nM and so the rate of loss of NADH can be measured spectrophotometrically (see Figure 2.5).

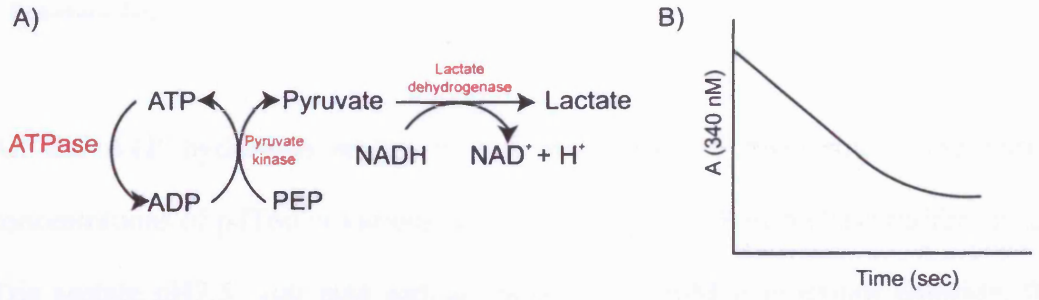

Figure 2.5. The coupled ATP-hydrolysis assay

A) The coupling of ATP hydrolysis to the oxidation of NADH in order to measure the rate of ATP hydrolysis.

B) A representation of a typical spectrophotometric trace of an assay from the spectrophotometer.

The rate of loss of NADH can be calculated using Equation 2.1.

$$dA_{340}=\varepsilon(d[NADH])l$$

Where dA_{340} = rate of change of absorbance UV light of 340 nM wavelength, ε = the absorbance coefficient of NADH, d[NADH] = rate of change of NADH concentration and l = path length of the spectrophotometer.

Equation 2.1.

As the rate of NADH oxidation to NAD^+ is equal to the rate of ATP hydrolysis to ADP by the ATPase, the concentration of ATP hydrolysed per second can be calculated using Equation 2.2.

$$dA340/6250 = d[ATP]\ M/sec$$

Where l = 1 cm, and $\varepsilon = 6250\ M^{-1}\ cm^{-1}$

Equation 2.2.

The rate of ATP hydrolysis, (v) (the moles of ATP hydrolysed per mole of ATPase per second) can be calculated using Equation 2.3.

$$d[ATP]/[ATPase] = v\ (s^{-1})$$

Equation 2.3.

All the ATP hydrolysis assays in this thesis used 5 nM protein and various concentrations of pdT60 or various concentrations of ATP in ATPase buffer (50 mM Tris acetate pH7.5, 100 mM sodium chloride, 10 mM magnesium chloride, 0.11 mg/ml NADH, 2 mM phosphoenolpyruvate, 20 U lactate dehydrogenase, 35 U pyruvate kinase). The reaction was initiated by the addition of protein and the A_{340} was measured using a spectrophotometer (UV 1601, Shimadzu).

The rate of ATP hydrolysis, v, was determined at various ATP concentrations ([S]) under saturating pdT60 concentrations. Values for k_{cat} (the turnover number), and K_{ATP} (the concentration of ATP when the rate of ATP hydrolysis, v, is equal to half the value of k_{cat}) were determined by fitting these data to the Michaelis-Menten

equation (Michaelis and Menten 1913) (see Equation 2.4). These data were fitted using Prism 4 (GraphPad software) by non-linear regression.

$$v = (V_{max}[S])/(K_m + [S])$$

Where $K_m = K_{ATP}$ and $V_{max} = k_{cat}[ATPase]_o$

Equation 2.4.

In order to determine K_{DNA}, the rate of ATP hydrolysis, v, was determined at various DNA concentrations ([S]) under saturating ATP concentrations. Values for k_{cat} (the turnover number), and K_{DNA} (the concentration of DNA when the rate of ATP hydrolysis, v, is equal to half the value of k_{cat}) were determined by fitting these data to the Michaelis-Menten equation as done for determining K_{ATP}

2.5.4 DNA-binding assay (gel-mobility shift assay)

2.5.4.1 Experimental setup

DNA-binding assays were carried out using protein diluted in binding buffer (20 mM Tris acetate pH7.5, 100 mM sodium chloride, 10 mM magnesium chloride, 0.1 mg/ml BSA, 10 % v/v glycerol, 0.1 mM DTT). The protein was incubated at room temperature for 5 minutes, radio-labelled DNA substrate was added and the mixture further incubated for 10 minutes. DNA binding was analysed using native polyacrylamide gels run in 1 x TBE buffer at 20 mA for 15 minutes at room temperature. The gels were dried and exposed to phosphor-imager screens (Molecular Dynamics). Quantitive analysis was done using a phosphor-imager and ImageQuant software (Molecular Dynamics). The amounts of free and bound DNA were quantified as counts of radioactivity for each lane. The % DNA substrate bound was calculated as being the number of counts of bound DNA substrate divided by

the total counts of both the bound and free DNA substrate. The % DNA substrate bound (Y) was plotted against protein concentration (X) and a value for K_d was obtained by fitting the data to Equation 2.5. These data were fitted using Prism 4 (GraphPad software) by non-linear regression.

$$Y = (Y_{max}X)/(K_d + X)$$

Where Y_{max} = 100 (% DNA bound) and K_d = X when Y =1/2 Y_{max}

Equation 2.5.

For all the DNA-binding assays 1 nM DNA substrate was used. For details of protein concentrations see figures in the results section.

2.5.4.2 Theory relating to the calculation of K_d

In the DNA binding assays the K_d is taken as equal to the protein concentration required to bind 50% of the DNA substrate. This can be equated to K_d as defined by:

$$K_d = \frac{[\text{Free Protein}][\text{Free DNA}]}{[\text{Protein/DNA complex}]}$$

If the DNA concentration used in the assays is significantly lower than the K_d of the protein for the DNA substrate, then the amount of protein in the protein/DNA complex will be negligible compared to the amount of free protein. Therefore [Free Protein] ≅ [Total Protein] meaning the K_d can be now described as:

$$K_d = \frac{[\text{Total Protein}][\text{Free DNA}]}{[\text{Protein/DNA complex}]}$$

When 50% of the DNA is bound in the protein/DNA complex, then:

$$\frac{[\text{Free DNA}]}{[\text{Protein/DNA complex}]} = 1$$

Therefore K_d = [Total Protein] when 50% of the DNA is bound in the protein/DNA complex. This means the K_d is equivalent to the amount of protein required to bind

50% of the DNA. This is not ideal and will not provide as accurate a K_d as isothermal calorimetry. This is because the assumption has to be made that the free DNA and protein/DNA complex is at equilibrium, but this equilibrium will be disturbed when the complex is separated from the free DNA and free protein in the electrophoresis gel. However, the assay is very useful to compare the relative DNA binding activities of different proteins under identical conditions.

2.5.5 Isothermal calorimetry (ITC)

In this thesis ITC was used to determine the affinity of protein-DNA interactions. All calorimetric experiments performed in this thesis were carried out using a Microcal VP-ITC instrument and all the experiments were performed as described in the manufacturers manual.

2.5.5.1 Sample preparation

The protein and DNA used in the experiment were dialysed into ITC buffer (50 mM Hepes pH 7.5, 100 mM sodium chloride and 10 mM magnesium chloride) overnight at 4°C. A change of buffer is necessary to remove DTT, Tris buffer and EDTA. DTT and Tris buffer can cause erroneous background signals due to the oxidation of DTT and the heat of ionisation of Tris buffer. EDTA chelates magnesium and will alter the magnesium concentration. Protein and DNA concentrations were determined spectrophotometrically and were degassed.

2.5.5.2 Experimental setup

The main part of the ITC instrument comprises two separate identical enclosed cells. One is a reference cell filled with double distilled, filtered and degassed water; the other is a reaction cell in which the two reactants are mixed. These two cells are

surrounded by two temperature-constant shields at the same temperature as the cells, in order to minimise heat exchange between the cells and the external surroundings.

In this thesis the protein was in the reaction cell and the DNA substrate was introduced into the reaction cell via a spinning syringe that homogenises the reactants. If there is an enthalpy change resulting from the interaction of the protein and the DNA substrate, the change in temperature of the reaction cell is detected, and energy is either supplied or removed from the reaction cell in order to take the temperature of the reaction cell back to that of the reference cell. The amount of energy that is supplied to the reaction cell is plotted against time. Further injections of the DNA substrate into the reaction cell occur and the amount of energy supplied to the cell measured and plotted. The integrated area under each peak is equal to the amount of heat evolved or absorbed during that injection. The integrated values of the peaks are plotted against time. A binding isotherm curve can be fitted by the least-squares method using the Origin software supplied with the ITC instrument and values for K_a and N are output. K_a is defined in Equation 2.6.

$$K_a = [\text{protein/DNA complex}]/[\text{protein}][\text{DNA}]$$

Equation 2.6.

The K_a can be used to calculate the inverse K_d. N can be used to calculate the number of molecules of protein bound to one DNA molecule, *N*, where $N=N^{-1}$. If, as explained in Chapter 4 (see section 4.3.2.1), two binding events with different equilibrium constants occur then a bipartite-isotherm binding curve is observed. The two parts of the curve each gives values for K_a and N, K_{a1} and N_1 for the first binding event and K_{a2} and N_2 for the second-binding event. In this case the number

of molecules of protein bound to one DNA molecule of the first and second binding events are N_1 and N_2 respectively, where $N_1 = N_1^{-1}$ and $N_2 = (N_1+N_2)^{-1}$.

2.5.5.3 Experimental parameters

The experimental parameters for all the ITC experiments are listed in Table 2.5.

Table 2.5. Experimental parameters for the ITC experiments

Parameter		Value
Shield, reference cell and reaction cell temperature		30 °C
Number of injections		30
Injection volume		1^{st} = 2 µl 2^{nd} – 30^{th} = 10 µl
Injection spacing		300 s
Syringe mix speed		310 rpm
Protein concentration		8 µM
DNA substrate concentration	pdT lengths 8-15	80 µM
	pdT lengths 20-30	50 µM

The first injection of the ITC experiment is a small volume of 2 µl to compensate for any losses or dilution-effects from the tip of the syringe during loading of the DNA into the syringe. Data from the first injection was removed during data analysis.

2.5.6 Helicase assay (strand-displacement assay)

Helicase assays were performed using the 5'-tailed duplex substrate described earlier in section 2.5.1.1.1. Helicase time course assays were carried out at 30 °C in helicase buffer (20 mM Tris acetate pH 7.5, 100 mM sodium chloride, 10 mM magnesium

chloride, 0.1 mg/ml BSA, 0.1 mM DTT) using 1 nM protein and 1 nM DNA substrate unless otherwise stated. The protein and the DNA substrate were incubated in the helicase buffer at 30 °C for 5 minutes before adding ATP to a final concentration of 1 mM to start the reaction. At each time point an aliquot was removed and the reaction was stopped with quenching buffer (0.2 % w/v SDS, 40 mM EDTA, 10 % w/v glycerol, 0.1% w/v bromophenol blue). DNA unwinding was analysed using SDS polyacrylamide gels run in SDS-PAGE buffer at 80 V for 60 minutes at room temperature. The gels were dried and exposed to phosphor-imager screens (Molecular Dynamics). Quantitive analysis was done using a phosphor-imager and ImageQuant software (Molecular Dynamics). The amount of duplex DNA and amount of ssDNA were determined for each lane and the % of the DNA substrate unwound was calculated. For the purposes of this thesis $t_{1/2}$ is the time taken by the protein to unwind 50 % of the DNA substrate when 1 nM protein and 1 nM DNA substrate are used.

The % DNA substrate unwound (Y) was plotted against time (X) and a value for $t_{1/2}$ was obtained by fitting the data to Equation 2.7. These data were fitted using Prism 4 (GraphPad software) by non-linear regression.

$$Y = (Y_{max}X)/(t_{1/2} + X)$$

Where Y_{max} = 100 (% DNA unwound) and $t_{1/2}$= X when Y =1/2 Y_{max}

Equation 2.7.

2.5.7 Processivity helicase assay

Processivity assays were performed using the 5'-tailed duplex substrate and the 40bp 5'-tailed duplex substrate described earlier in section 2.5.1.1.1

The protein and DNA substrate were incubated in processivity buffer (20 mM Tris acetate pH7.5, 100 mM sodium chloride, 10 mM magnesium chloride, 0.1 mg/ml BSA, 0.1 mM DTT) using 1 μM protein and 1 nM DNA substrate. The reaction was started by the addition of start buffer (1 mM ATP, 50 μM heparin (final concentrations)). The assays were analysed as described for the helicase assay earlier in section 2.5.6.

2.6 X-ray crystallography methods

2.6.1 DNA substrate S1

DNA S1 comprised oligonucleotide 6 (O6) and oligonucleotide 7 (O7) forming a 5'-tailed duplex DNA (see Figure 2.6). The duplex DNA was 10 base pairs and the 5' tail was pdT7.

```
O6 3'          CGTCACGAGC 5'
O7 5' TTTTTTTGCAGTGCTCG 3'
```

Figure 2.6. DNA S1

DNA S1 was purified, after annealing of O6 and O7, on a resource Q column. DNA S1 in annealing buffer (50 mM phosphate buffer pH 6.8, 100 mM potassium chloride, 10 mM magnesium chloride) was loaded onto a resource Q column equilibrated with 50 mM phosphate buffer pH 6.8 , 100 mM potassium chloride. After washing away unbound material DNA S1 was eluted with a 30 column-volume linear gradient of 0.05 – 1 M potassium chloride. DNA S1 was exchanged into water using a NAP25 drip column, concentrated using a centricon (Micon) with an appropriate molecular weight cut-off and stored at -20 °C.

2.6.2 Crystallisation

Crystallisation trials for protein and protein/DNA complex crystals were set up using the high-throughput Mosquito robot (TTP Lab-Tech) using the sitting drop method (100 nl drop size), commercially available sparse matrix screens and 96-well sitting-drop high-throughput plates (Axygem). Crystallisation trials were also set up by hand using the hanging drop vapour diffusion method (2 µl drop size) and sparse matrix screens in 24 well VDX plates (Hampton Research) with plastic cover slips (Henleysmed). The screens used are listed below in Table 2.6.

Table 2.6. Crystallisation screens

Supplier	**High-throughput screens**	**Screens set up by hand**
Hampton Research	Crystal screen HT Index screen HT Salt RX screen HT	Crystal screen Crystal screen 2 Natrix
Molecular Dimensions	Clear strategy screen I Clear strategy screen II Structure screen I & II Memsys & Memstart	
Jena Biosciences	JBScreen classic I JBScreen classic II	
Emerald Biosciences	Wizard screen HT	

Potential crystallisation conditions were optimised using grid screens around the lead condition. Protein/DNA complex crystals were grown in 2 µl hanging drops using

the vapour diffusion method in 24 well VDX plates (Hampton Research) with plastic cover slips (Henleysmed).

2.6.2.1 Micro-seeding

Protein/DNA complex crystals were crushed using a glass pipette, the end of which had been melted into a ball shape, or by using a crystal-probe (Hampton Research). The resulting solution was diluted by 10, 100 and 1000 in the mother liquor and 0.5 µl of a diluted stock was added to a 2 µl hanging drop.

2.6.3 Data collection

Data was collected from crystals using either the in-house copper rotating anode X-ray source (Rigaku) with a MAR 345 scanner (MAR Research), or at the European Syncotron Radiation Facility (ESRF) in Grenoble, or at the Syncotron Radiation Source (SRS) in Daresbury. For all data collections, 1° of data was collected at 0°, 45° and 90° of phi. The images were autoindexed using MOSFLM (CCP4 suite (CCP4 1994)) and the space group and unit cell were determined. The strategy option of MOSFLM was used to determine the degrees of Phi to collect the dataset with the highest completeness.

2.6.3.1 Room temperature

Crystals were mounted in quartz capillaries between layers of mother liquor. The capillary ends were sealed with wax. The capillary was mounted onto a goniometer head for data collection.

2.6.3.2 100 °K

Crystals were cryoprotected using a sequential soaking method. Crystals were placed in 5 µl 1.1 x mother liquor. 5 µl 1.1 x mother liquor with 10% MPD was added to

the drop and was mixed before removing 5 µl, leaving the crystal in 5 µl of 1.1 x mother liquor with 5 % MPD. This was repeated by adding 5 µl 1.1 x mother liquor with 15 % MPD, leaving the crystal in 5 µl 1.1 x mother liquor with 10% MPD. This process was repeated with 5 % increments of MPD until the crystal was in 1.1 x mother liquor with 20 % MPD. Crystals were flash frozen in the cryo-stream (Oxford Cryosystems) of the in-house X-ray set.

2.6.4 Data processing

All datasets were processed using default settings except where stated.

2.6.4.1 Integration and scaling

All the data sets of the short c-axis were processed using MOSFLM. The space group and unit cell were determined as described earlier in this section. The unit cell was refined and the mosaicity determined using four 5° sections of data 30° apart. The whole dataset was then integrated as a single output reflection file. The data was scaled using SCALA (CCP4 suite (CCP4 1994)). Data was merged during scaling except for derivative datasets where Friedel pairs were kept separate.

All the datasets collected from crystals with the long c-axis were processed using XDS (Kabsch, W). The space group and unit cell were determined using three 10° sections of data 30° apart. The whole dataset was then integrated using these parameters. XDS provides an un-scaled statistical output that can be used to check the quality of the processed data, as well as a refined unit cell. The refined unit cell was then used to reprocess the whole dataset and the process repeated until the refined unit cell matched the input unit cell. The data was then scaled into a single

data file using XSCALE (Kabsch, W). Data was merged during scaling except for derivative datasets where Friedel pairs were kept separate.

2.6.4.2 Phase determination

Native and derivative datasets were scaled together using SCALEIT (CCP4 suite (CCP4 1994)).

Heavy atom positions were searched for using the automatic functions of the auto.SHARP and SOLVE programmes (Bricogne *et al.* 2003; Terwilliger and Berendzen 1999). Difference Patterson maps were calculate using the FFT programme (CCP4 suite (CCP4 1994)) and peaks searched for manually.

Potential heavy atom positions were refined in SHARP (Bricogne *et al.* 2003), which also calculated experimental electron density maps using the resulting phasing information. Coherent and protein-like electron density was then looked for manually.

2.7 Buffer formulations

2.7.1 Cell culture

Luria broth (LB)	As described by Sambrook *et al.* (Sambrook *et al.* 1989). For plating purposes 0.01 g/ml agar was added
Seleno-methionine media (per litre)	0.5 g alanine, 0.58 g arginine HCl, 0.41 g aspartic acid, 0.03 g cystine 0.67 g (0.75 g) glutamic acid (Na) 0.33 g glutamine, 0.54 g glycine 0.06 g histidine 0.23 g isoleucine 0.23 g leucine 0.23, 0.42 g lysine HCl 0.13 g phenylalanine, 0.10 g proline 2.08 g serine, 0.24 g threonine 0.17 g tyrosine, 0.23 g valine 0.50 g adenine, 0.67 g guanosine 0.17 g thymine, 0.50 g uracil 1.50 g sodium acetate, 1.50 g succinic acid 0. 75 g ammonium chloride 1.08 g Sodium hydroxide 8.0 g (10.5 g) di-potassium hydrogen orthophostphate (trihydrate)
10 x seleno-methionine media supplement (per 200 ml) Added to sterile seleno-methionine media before growing culture	20g glucose 0.5g magnesium sulphate 7-hydrate 8.35mg iron sulphate 16.6ul conc sulphuric acid 10 mg thiamine 100 mg seleno-methionine

Where appropriate 100 μg/ml ampicillin, 34 μg/ml chloramphenicol and 100 μg/ml spectinomycin were included (final concentrations).

2.7.2 Gel electrophoresis

2.7.2.1 Loading buffers

SDS-PAGE loading buffer (5x)	7% v/v BME 10% w/v SDS 0.5 % w/v bromophenol blue 20% glycerol
Agarose gel loading buffer (10x)	0.25 % w/v bromophenol blue 50% w/v sucrose 10 mM Tris/HCl pH 8.0

2.7.2.2 Running buffers

SDS-PAGE buffer (5x stock), (per litre)	72g glycine 15 g Tris base 5.0 g SDS
TBE (5x) (per litre)	54.0 g Tris/HCl base 27.5 g boric acid 20 ml 0.5 M EDTA

2.7.2.3 Staining/destaining

Coommassie brilliant blue stain.	25 % v/v methanol 8 % v/v acetic acid 0.2 % w/v Coomassie brilliant blue
Destain	25 % v/v methanol 8% v/v acetic acid
Ethidium bromide stain	0.1 mg/ml in water

2.7.3 Protein purification

D1	20 mM Tris/HCl pH 7.5 500 mM sodium chloride 100 μM phenyl methyl sulfonyl fluoride 10 μM leupeptin 1 μM pepstatin
D2	20 mM Tris/HCl pH 7.5 40 mM imizadole
D3	20 mM Tris/HCl pH 7.5 500 mM sodium chloride
D4	20 mM Tris/HCl pH 7.5 1 mM EDTA 1 mM DTT
D5	20 mM Tris/HCl pH 7.5 100 mM sodium chloride 1 mM EDTA 1 mM DTT
I1	20 mM Tris/HCl pH 7.0 500 mM sodium chloride 1 mM EDTA
I2	20 mM Tris/HCl pH 7.0 500 mM sodium chloride 50 mM MESNA 1 mM EDTA
I3	20 mM Tris/HCl pH8.5 500 mM sodium chloride 1 mM EDTA
I4	20 mM Tris/HCl pH 8.0 500 mM sodium chloride 1% (w/v) SDS

3 An attempt to make *Bacillus stearothermophilus* PcrA mutants and an *Archaeoglobus fulgidus* PCNA dimer using the IMPACT-TWIN intein system

3.1 Introduction

3.1.1 Summary

This chapter describes the attempt to create mutants of the *Bacillus stearothermophilus* PcrA (*bs*PcrA) helicase and to create a dimer of the *Archaeoglobus fulgidus* proliferating cell nuclear antigen (*af*PCNA). For both of these projects modified inteins were to be used to produce the required final protein products. Therefore this chapter first explains how the modified inteins work and then goes on to describe the work done with *bs*PcrA and *af*PCNA.

3.1.2 Inteins

Protein expression and purification can often be tricky with many problems to overcome. A given protein may be toxic to the host organism. At other times it may be desirable to make a fusion protein, but either it cannot be cloned or will not express. Inteins are proteins that have the ability to form reactive polypeptides that can then be ligated together. This chapter describes efforts to use inteins to express a toxic protein, a *bs*PcrA mutant, and a covalently linked dimer of *af*PCNA.

Inteins are naturally occurring protein elements that have the ability to splice themselves out of proteins. The IMPACT-TWIN system is an expression kit that fuses a modified intein and chitin-binding domain (CBD) to a target protein (New

England Biolabs; Watanabe *et al.* 1994). The resulting fusion protein can be purified by affinity chromatography using a chitin column. The target protein is eluted by inducing a specific intermolecular proteolysis by the intein that results in the target protein being cleaved off the column-bound intein.

The IMPACT-TWIN system can also be used to create two reactive polypeptides for intein-mediated protein ligation (IPL) (see Figures 1.1-1.3) (Dawson *et al.* 1994; Evans *et al.* 1998, 1999a, b; Evans and Xu 1999; Mathys *et al.* 1999; Muir *et al.* 1998; Severinov and Muir 1998; Tam *et al.* 1995; Xu *et al.* 1999). This technique allows the ligation of two polypeptides *in vitro*, which is useful for making proteins that cannot be made genetically. For example proteins that are toxic *in vivo* can be over expressed as two non toxic fragments and ligated together *in vitro*. If a protein comprising two domains, P1 and P2, is toxic to *E.coli* when expressed as full length protein, the two domains can be prepared separately and then ligated together *in vitro*. P1 is cloned into either the pTWIN1 or the pTWIN2 expression plasmid in such a way that it is fused via its C-terminus to a modified intein derived from either the *Mycobacterium xenopi* GyrA gene (*Mxe*GyrAi) or from the *Methanobacterium thermoautotrophicum* RIR1 gene (*Mth*RIR1i) (Evans *et al.* 1998, 1999b; Iwai and Pluckthun 1999; Telenti *et al.* 1997) (see Figure 3.1A). P2 is cloned into pTWIN1 in such a way that it is fused via its N-terminus to a modified intein derived from the *Synechocystis species* DnaB gene (*Ssp*DnaBi) (Mathys *et al.* 1999; Wu *et al.* 1998) (see Figure 3.1B).

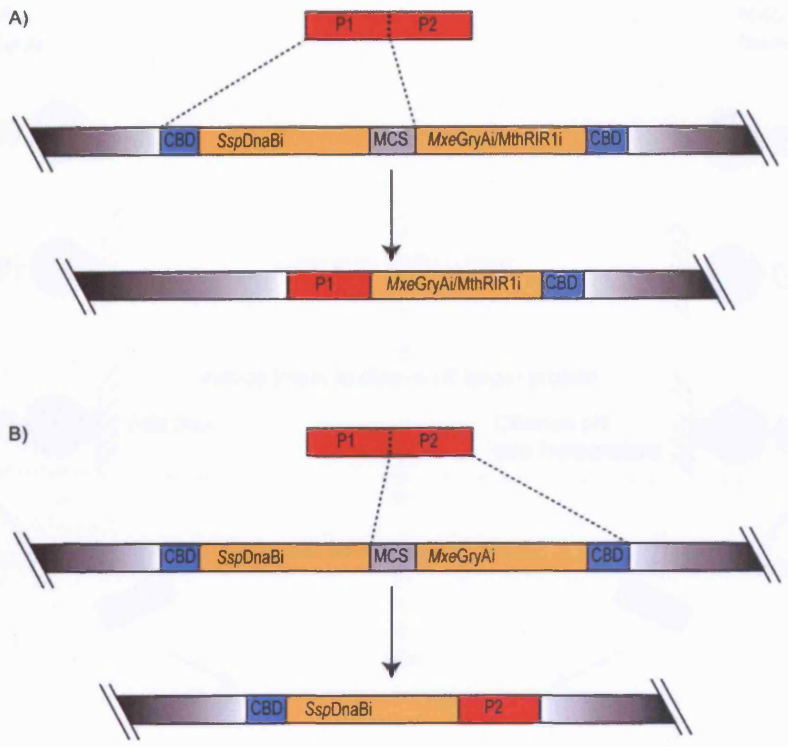

Figure 3.1. Cloning P1 and P2 into pTWIN1/pTWIN2

A) The P1 gene fragment is fused via its C-terminus to the *Mxe*GyrAi mini intein (*Mxe*GyrAi) in pTWIN1, or *Mth*RIR1 mini intein in pTWIN2 (*Mth*RIR1i).

B) The P2 gene fragment is fused via its N-terminus to the *Ssp*DnaBi mini intein (*Ssp*DnaBi) in pTWIN1.

The only requirement for intein mediated protein ligation is that the N-terminal residue of P2, the residue to be ligated, must be a cysteine residue and the C-terminal residue of P1 must have a thiol group attached (Evans *et al.* 1998; New England Biolabs). P1 is cleaved from the chitin-bound intein by addition of the thiol 2-mercaptoethanesulfonic acid (MESNA), allowing the purification of P1 with a C-terminal thiol ester (see Figure 3.2). P2 is cleaved from the chitin-bound intein by lowering the pH and raising the temperature allowing the purification of P2 with an N-terminal cysteine (see Figure 3.2).

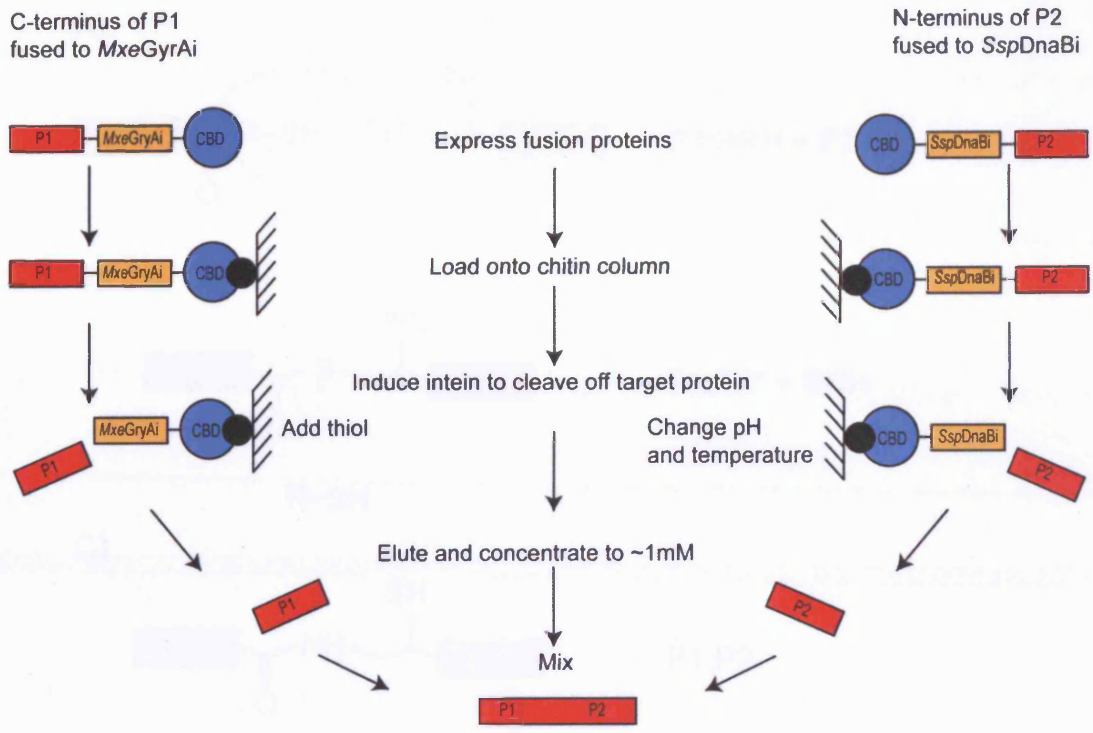

Figure 3.2. Strategy to purify and ligate P1 and P2 protein fragments to make the P1-P2 fusion protein

A)The P1-*Mxe*GryAi fusion protein is loaded onto a chitin column at pH 7.0 and 4 °C. The *Mxe*GryAi is induced to cleave off the P1 by the addition of a thiol (MESNA) and P1 is eluted from the column.
B)The *Ssp*DnaBi-P2 fusion protein is loaded onto a chitin column at pH 8.5 and 4 °C. The *Ssp*DnaBi is induced to cleave off P2 by lowering the pH to 7.0 and raising the temperature to 22 °C and P2 is eluted from the column.
C) The P1 and P2 fragments are concentrated and mixed allowing them to ligate together to form the final P1-P2 protein.

P1 and P2 are concentrated and mixed allowing the thiol of P1 to react with the cysteine of P2 to form a peptide bond between P1 and P2 (see Figure 3.3A, B and C) (Dawson *et al.* 1994; Tam *et al.* 1995). The thiol attached to the C-terminal of P1 can dissociate in a reversible fashion, but this does not effect the rate of formation of P1:P2, which is independent of the free thiol concentration (see Figure 3.3 D and Equation 3.7.).

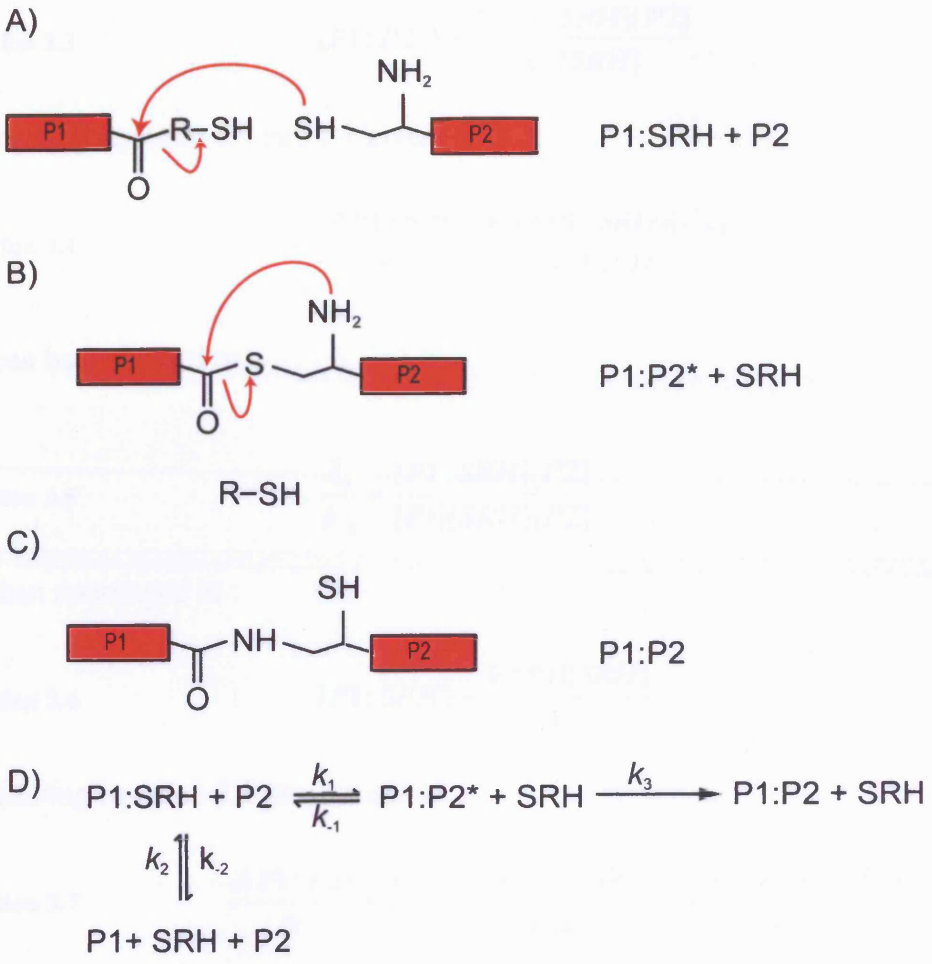

Figure 3.3. Peptide bond formation between P1 and P2

A) A reversible thiol/thiol ester exchange gives a thiolester linked intermediate. B) An irreversible nucleophilic rearrangement occurs resulting in C) A peptide bond between P1 and P2. D) The overall reaction.

Parts A), B) and C) are based on a figure from (New England Biolabs)

The overall rate of formation of P1:P2 can be expressed as:

Equation 3.1

$$\frac{d[P1:P2]}{dt} = k_3[P1:P2^*]$$

$\frac{k_1}{k_{-1}}$ can be expressed as:

Equation 3.2

$$\frac{k_1}{k_{-1}} = \frac{[P1:P2^*][SRH]}{[P1:SRH][P2]}$$

Or when re-arranged as:

Equation 3.3

$$[P1:P2^*] = \frac{k_1[P1:SRH][P2]}{k_{-1}[SRH]}$$

Substituting equation 1.3 into 1.1 gives:

Equation 3.4

$$\frac{d[P1:P2]}{dt} = \frac{k_3 k_1[P1:SRH][P2]}{k_{-1}[SRH]}$$

$\frac{k_2}{k_{-2}}$ can be expressed as :

Equation 3.5

$$\frac{k_2}{k_{-2}} = \frac{[P1:SRH][P2]}{[P1][SRH][P2]}$$

Or when rearranged as :

Equation 3.6

$$[P1:SRH] = \frac{k_2[P1][SRH]}{k_{-2}}$$

Substituting equation 3.6 into 3.4 gives:

Equation 3.7

$$\frac{d[P1:P2]}{dt} = \frac{k_3 k_1[P2]k_2[P1][SRH]}{k_{-1}[SRH]k_{-2}} \Rightarrow \frac{k_1 k_2 k_3[P2][P1]}{k_{-1}k_{-2}}$$

Equation 3.7 demonstrates that the rate of formation of P1:P2 is not dependent on the concentration of the thiol SRH.

3.2 An attempt to make full length *Bacillus stearothermophilus* PcrA from two halves

3.2.1 Introduction

Bacillus stearothermophilus PcrA (*bs*PcrA) is a Superfamily one (SF1) 3'-5' helicase involved in DNA repair and rolling cycle plasmid replication (Petit *et al.* 1998; Soultanas *et al.* 1999). The crystal structure of *bs*PcrA shows that it has two domains each of which comprises two sub domains (Velankar *et al.* 1999) (see Figure 3.4). For the purposes of this work domains 1A and 1B are referred to as *bs*PcrA1 and domains 2A and 2B are referred to as *bs*PcrA2 (see Figure 3.4).

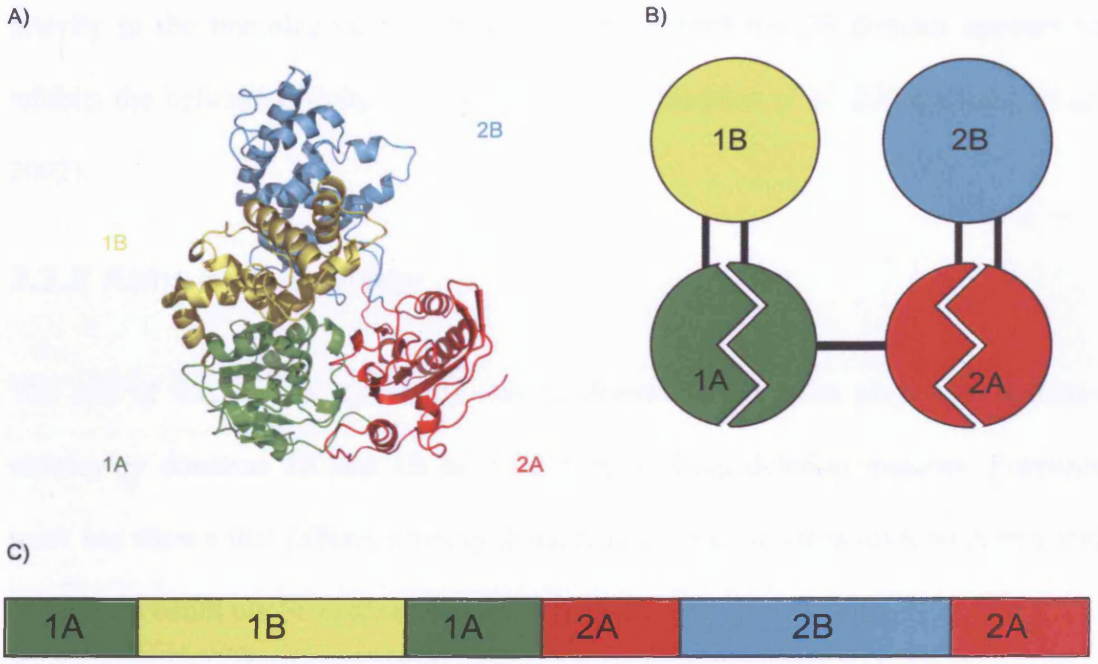

Figure 3.4. The structure of *bs*PcrA

The domains of *bs*PcrA are coloured as follows: 1A-green, 1B-yellow, 2B-blue, 2A-red. *bs*PcrA1 comprises domains 1A and 1B. *bs*PcrA2 comprises domains 2A and 2B.

A) The crystal structure of *bs*PcrA. B) A schematic representation of the crystal structure of *bs*PcrA. C) A schematic of the *bs*PcrA gene.

Like all SF1 helicases, *bs*PcrA has the seven classical helicase motifs (Gorbalenya and Koonin 1993). Crystal structures of *bs*PcrA in complex with DNA with ATP analogues have been used to propose an inchworm model of helicase activity (Velankar *et al.* 1999). Velankar *et al.* propose that domains 1A and 2A (the motor domains) translocate along ssDNA and domains 1B and 2B interact with the duplex DNA (via amino acids K456, T426, K419 of 2B and K137, K138 of 1B) ahead of the motor domains (Velankar *et al.* 1999). This interaction disrupts the structure of the duplex DNA allowing separation of the two strands (Velankar *et al.* 1999). This hypothesis is supported by biochemical evidence in that helicase activity is reduced when the amino acids involved with duplex interaction are mutated to alanine (Soultanas *et al.* 2000). In contrast, the 2B domain is not required for helicase activity in the homologous SF1 helicase Rep, instead the 2B domain appears to inhibits the helicase activity of monomeric Rep (Brendza *et al.* 2005; Cheng *et al.* 2002).

3.2.2 Aims and strategy

The aim of this part of the thesis was to determine the roles played in helicase activity by domains 2B and 1B of *bs*PcrA by making deletion mutants. Previous work has shown that *bs*PcrA missing domain(s) 2B and/or 1B is toxic to *E.coli* and the protein could not be expressed (personal communication, Wigley, D.).

The IMACT-TWIN system was used to try and create the deletion mutants. Figure 3.5 outlines the strategy used. The aim was to separately purify two halves of *bs*PcrA, with or without domains 2B and 1B, and then ligate them together *in vitro* to form full-length *bs*PcrA or *bs*PcrA deletion mutants.

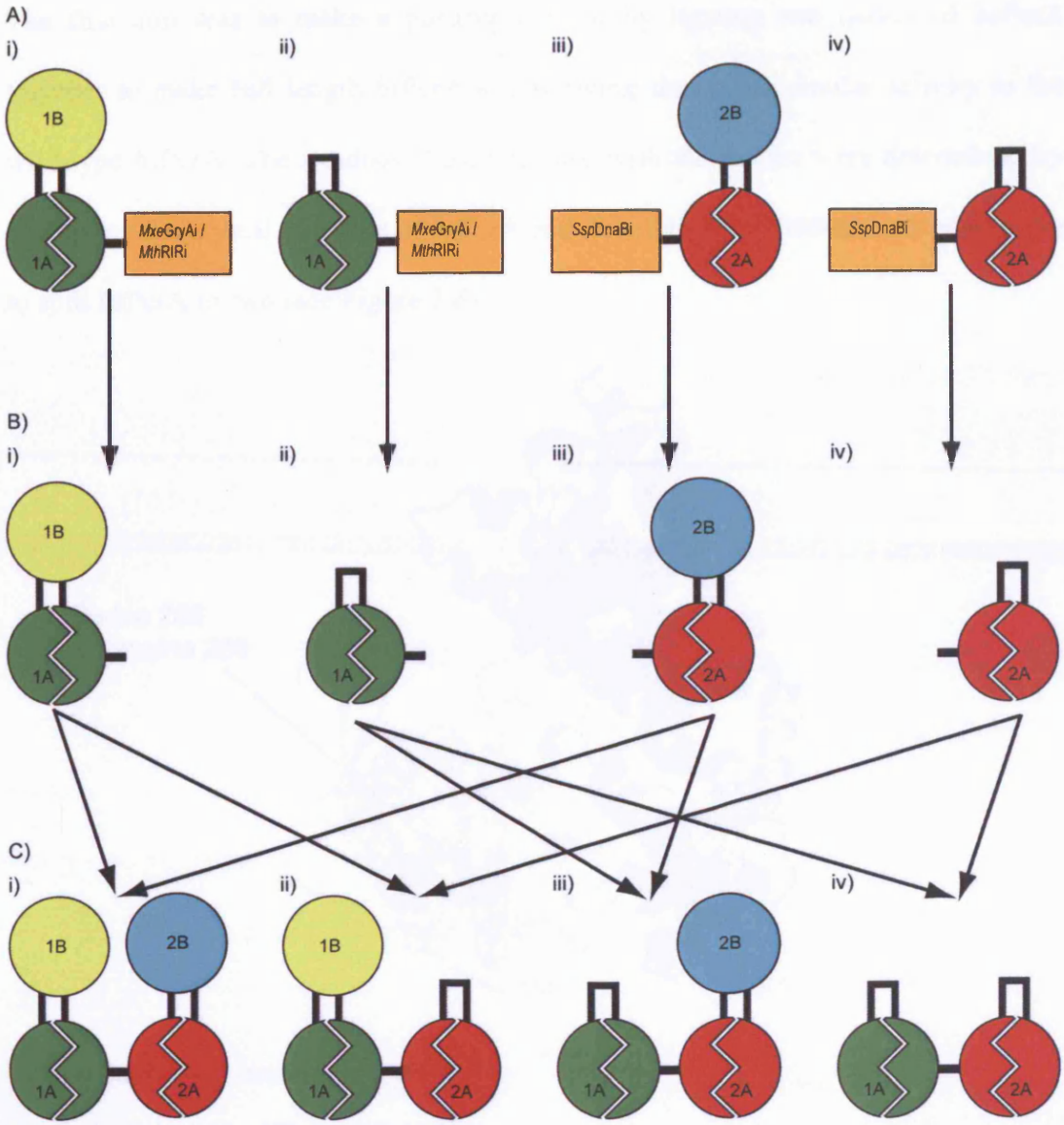

Figure 3.5. Strategy for creating *bs*PcrA deletion mutants

A) Fusion proteins to be made:

i) *bs*PcrA1-*Mxe*GryAi/*Mth*RIR1i,

ii) *bs*PcrA 1A-*Mxe*GryAi/*Mth*RIR1i,

iii) *Ssp*DnaBi *bs*PcrA2,

iv) *Ssp*DnaBi–*bs*PcrA 2A.

B) *bs*PcrA fragments after cleavage from inteins ready for ligation:

i) *bs*PcrA1,

ii) *bs*PcrA 1A,

iii) *bs*PcrA2,

iv) *bs*PcrA 2A.

C) Products of ligated *bs*PcrA fragments:

i) *bs*PcrA1 + *bs*PcrA2 = *bs*PcrA,

ii) *bs*PcrA1 + *bs*PcrA 2A = *bs*PcrA Δ2B,

iii) *bs*PcrA 1A + *bs*PcrA 2 = *bs*PcrA Δ1B,

iv) *bs*PcrA 1A + *bs*PcrA 2A = *bs*PcrA Δ1B2B.

The first aim was to make a positive control by ligating two halves of *bs*PcrA together to make full length *bs*PcrA and showing that it has similar activity to the wild-type *bs*PcrA. The residues chosen to fuse with the inteins were determined by examining the crystal structure of *bs*PcrA and deciding where would be a good point to split *bs*PcrA in two (see Figure 3.6).

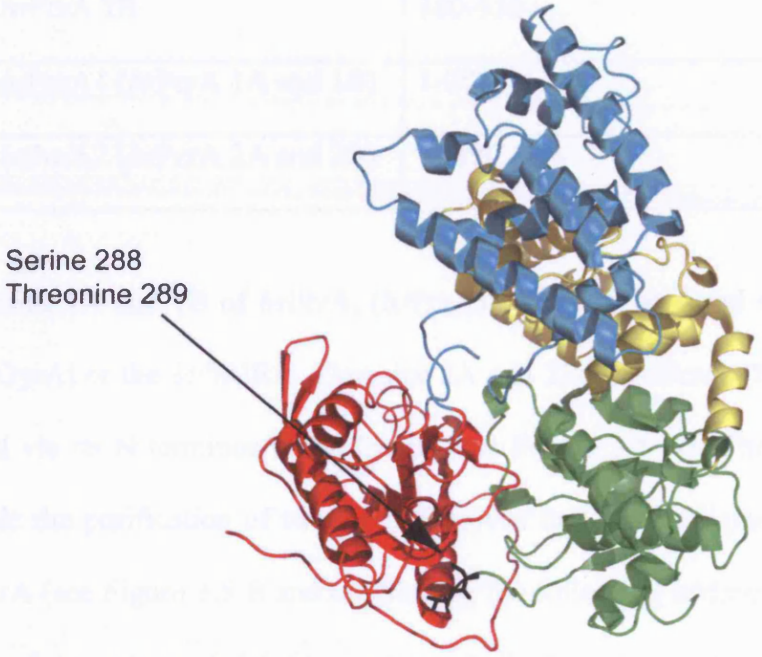

Figure 3.6. Crystal structure of *bs*PcrA

Serine 288 and threonine 289 are labelled in black.

The cut point decided upon was between residues serine 288 and threonine 289, the reason being that these residues are located in a flexible linker that connects the two naturally occurring halves of *bs*PcrA meaning that the secondary structure would be left intact and allow correct folding of the *bs*PcrA halves when they are fused to the modified inteins (see Figure 3.6). In addition intein mediated protein ligation requires a cysteine as the N-terminal residue of the protein and threonine was thought to be a suitable amino acid to be mutated to cysteine without changing any properties of the *bs*PcrA. The domains of *bs*PcrA are classified as listed below in Table 3.1.

Table 3.1. Definitions of *bs*PcrA domains

Domain of *bs*PcrA	Amino acids
*bs*PcrA 1A	1-90 and 219-288
*bs*PcrA 1B	91-218
*bs*PcrA 2A	289-379 and 550-724
*bs*PcrA 2B	380-550
*bs*PcrA1 (*bs*PcrA 1A and 1B)	1-288
*bs*PcrA2 (*bs*PcrA 2A and 2B)	T289C-724

Domains 1A and 1B of *bs*PcrA, (*bs*PcrA1), were to be fused via its C-terminus to *Mxe*GyrAi or the *Mth*RIR1i. Domains 2A and 2B of *bs*PcrA, (*bs*PcrA2), were to be fused via its N-terminus to *Ssp*DnaBi (see Figure 3.5 A). These constructs would enable the purification of two *bs*PcrA halves that can be ligated together to create *bs*PcrA (see Figure 3.5 B and C). Making the following additional constructs would allow the creation of deletion mutants: *bs*PcrA domain 1A (*bs*PcrA 1A) fused to *Mxe*GyrAi/ *Mth*RIR1i intein and *bs*PcrA domain 2A (*bs*PcrA 2A) fused to the *Ssp*DnaBi intein (see Figure 3.5 A). This would allow the following ligations: i) *bs*PcrA 1A and *bs*PcrA2 can create *bs*PcrA missing domain 1B (*bs*PcrA Δ1B), ii) *bs*PcrA1 and *bs*PcrA 2A can create *bs*PcrA missing domain 2B (*bs*PcrA Δ2B), iii) *bs*PcrA 1A and *bs*PcrA 2A can create *bs*PcrA missing domains 1B and 2B (*bs*PcrA Δ1B2B) (see Figure 3.5 B and C).

3.2.3 Cloning and expression of *bs*PcrA1-*Mxe*GyrAi and *bs*PcrA1-*Mth*RIR1i

The *bs*PcrA1 encoding sequence was cloned into both pTWIN1 and pTWIN2 to produce a construct with *bs*PcrA amino acids 1-288 cloned directly next to the

*Mxe*GyrAi or the *Mth*RIR1i intein respectively. Both the *bs*PcrA1-*Mxe*GyrAi and *bs*PcrA1-*Mth*RIR1i fusion proteins expressed in BL21 codon 2+ *E.coli*, but were insoluble. Expression trials were carried out, to try and reduce the level of expression to see if any soluble protein could be made. The expression strains BL21 codon 2+, Tuner codon 2+, BL21 codon 2+ pLysS, and were tested at either 3 hours for 37 °C or overnight at 18 °C with varying amounts of IPTG. The expression trials revealed that the *bs*PcrA1-*Mxe*GyrAi fusion protein was soluble when expressed in Tuner codon 2+ *E.coli* cells when induced with 0.2 mM IPTG for 18 hours at 18 °C, however this condition did not produce soluble *bs*PcrA1-*Mth*RIR1i fusion protein (see Figure 3.7).

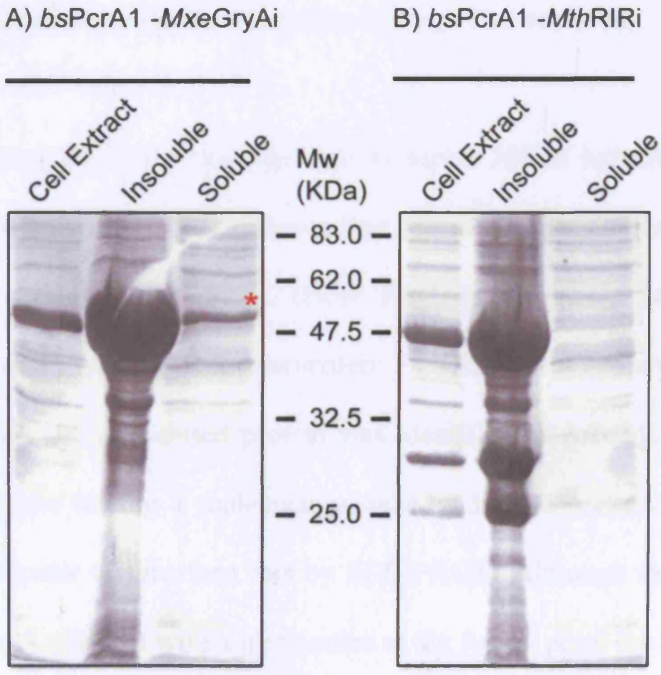

Figure 3.7. Expression of *bs*PcrA1-*Mxe*GryAi and *bs*PcrA1-*Mth*RIR1i

A) Expression of soluble *bs*PcrA1-*Mxe*GyrAi fusion protein showing the whole cell extract and insoluble and soluble fractions. * indicates soluble *bs*PcrA- *Mxe*GyrAi protein.

B) Expression of insoluble *bs*PcrA1-*Mth*RIR1i fusion protein showing the whole cell extract, and insoluble and soluble fractions.

3.2.4 Purification of *bs*PcrA1 from *bs*PcrA1-*Mxe*GyrAi

Small-scale purification tests were carried out in which the expressed *bs*PcrA1-*Mxe*GryAi fusion product was incubated with chitin beads. The beads were then pelleted, pulling down the protein-intein fusion and the supernatant discarded. The chitin beads were then re-suspended in a buffer designed to induce intein cleavage after which the chitin beads were pelleted to leave the purified protein in the supernatant. However, no *bs*PcrA1 was purified from *bs*PcrA1-*Mxe*GyrAi (see Figure 3.8 A). The chitin beads were then washed in an SDS solution to dissociate anything bound to the chitin beads and the full length *bs*PcrA1-*Mxe*GryAi fusion protein was eluted showing that no intein cleavage had occurred.

In order to stimulate intein cleavage activity serine 288 of *bs*PcrA1 was mutated to methionine because it has been shown that the intein cleaves more readily with a methionine at the fusion point (New England Biolabs). In the small-scale purification tests a protein of approximately 27 KDa was eluted after intein cleavage (see Figure 3.8 B). The eluted protein was identified as *bs*PcrA1 by peptide mass mapping. Despite having a molecular weight of 33KDa *bs*PcrA1 runs at a lower apparent molecular weight than this by SDS-PAGE. Although the cleavage by the intein was more efficient with a methionine at the fusion point the SDS wash reveals that not all the cleaved *bs*PcrA1 protein was soluble and insoluble *bs*PcrA1 protein was pelleted with the chitin beads after intein cleavage (see figure 1.8 B).

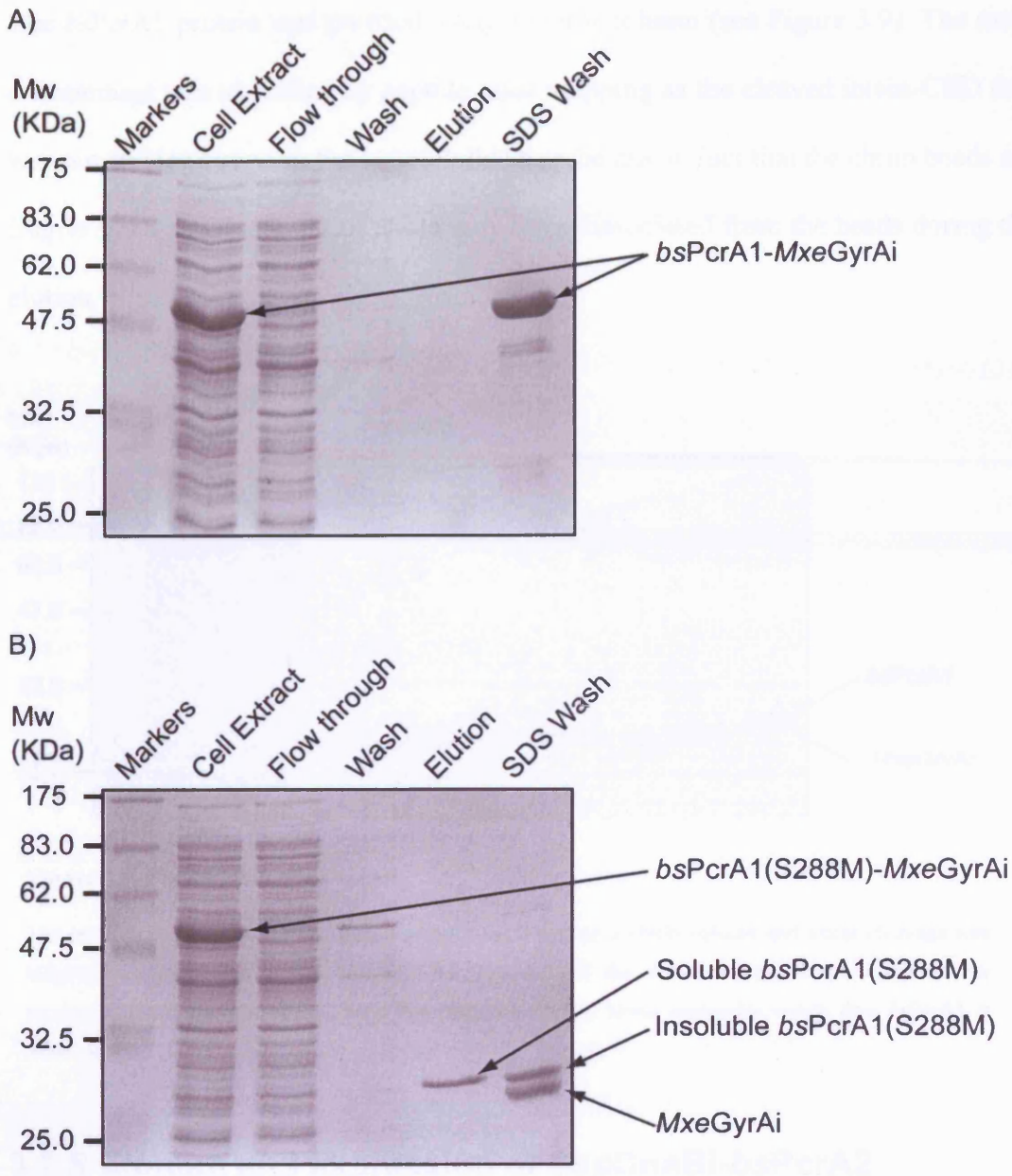

Figure 3.8. Small scale purification tests of *bs*PcrA1

A) *bs*PcrA1-*Mxe*GryAi. No *bs*PcrA1 protein is eluted from the chitin beads because the *bs*PcrA1-*Mxe*GryAi fusion protein remains intact.

B) *bs*PcrA1(S288M) – *Mxe*GryAi. The *bs*PcrA1 is eluted from the chitin beads and intein cleavage occurs leaving no full-length fusion protein intact. Some of the *bs*PcrA1 is insoluble and elutes only after washing the chitin beads in SDS.

The *bs*PcrA1 protein was purified using a chitin column (see Figure 3.9). The main contaminant was identified by peptide mass mapping as the cleaved intein-CBD that was not staying bound to the column; this may be due to fact that the chitin beads are fragile and a small amount of chitin may have dissociated from the beads during the elution.

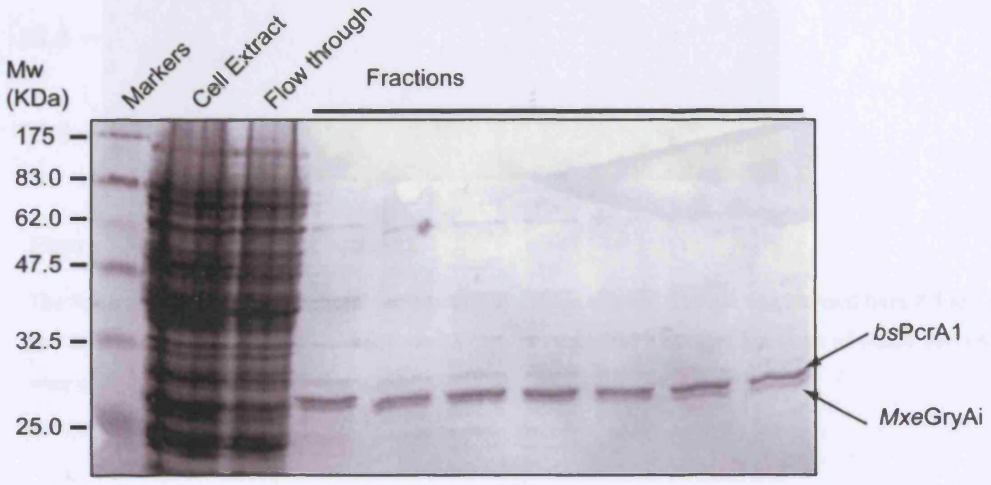

Figure 3.9. Purification of *bs*PcrA1

The *bs*PcrA1(S288M)-*Mxe*GryAi fusion protein was bound to a chitin column and intein cleavage was induced by the addition of the thiol MESNA. Fractions of the eluted *bs*PcrA1 were collected after incubation overnight. The second band that runs at a slightly lower molecular weight than *bs*PcrA1 is *Mxe*GryAi.

3.2.5 Cloning and expression of *Ssp*DnaBi-*bs*PcrA2

The sequence encoding for *bs*PcrA2, amino acids T289C-724, was cloned into pTWIN1 and expressed directly fused to the *Ssp*DnaBi intein via its N-terminus. Threonine 289 of *bs*PcrA was mutated to a cysteine as required for IPL.

3.2.6 Purification of *bs*PcrA2 from *Ssp*DnaB-*bs*PcrA2

The intein cleavage was efficient and the *bs*PcrA2 was purified from the chitin column (see Figure 3.10).

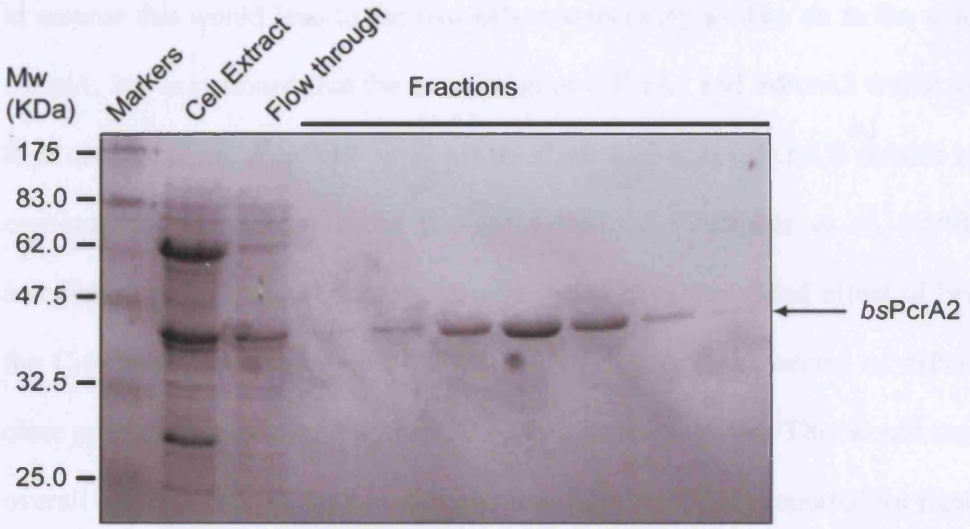

Figure 3.10. Purification of *bs*PcrA2

The *Ssp*DnaBi-*bs*PcrA2 fusion protein was bound to a chitin column. The pH was lowered from 8.5 to 7.0 and the temperature increased from 4°C to 22°C to induce intein cleavage. Fractions of eluted *bs*PcrA2 were collected after incubation overnight.

3.2.7 Ligation of *bs*PcrA1 and *bs*PcrA2 to form *bs*PcrA

For successful ligation of short peptides one of the peptides needs to be at a concentration of 1 mM. At this concentration most proteins would precipitate and become insoluble. The reports that 1 mM concentrations are needed come from the ligation of synthetic peptides, which can be concentrated to 1 mM, being ligated to proteins. Other reports have shown that concentrations of 0.8 mM are high enough for the ligation of two large proteins (Evans *et al.* 1999b). The high concentrations increase the chance of the C-terminal thiol ester and the N-terminal cysteine being in close enough proximity to allow the ligation reaction to proceed.

Some *bs*PcrA1 became insoluble after being cleaved from *Mxe*GryAi suggesting that *bs*PcrA1 in isolation may precipitate when concentrated. In order to stop *bs*PcrA1 precipitating, *bs*PcrA1 and *bs*PcrA2 were concentrated together as it was reasonable

to assume this would lead to the two halves associating as they do in the wild type *bs*PcrA. It was reasoned that the association of *bs*PcrA1 and *bs*PcrA2 would allow a high concentration of *bs*PcrA to be achieved, as wild-type *bs*PcrA is soluble at high concentrations as shown when it was crystallised (Velankar *et al.* 1999). The association of *bs*PcrA1 and *bs*PcrA2 would also have the added effect of bringing the C-terminal methionine of *bs*PcrA1 and the N-terminal cysteine of *bs*PcrA2 in close proximity increasing the chances of the ligation reaction. This would mean the overall concentration of protein may not need to be as high as reported for ligation to occur.

The *bs*PcrA1 and *bs*PcrA2 proteins were concentrated together. There was no need to buffer exchange because the reaction proceeds at neutral pH (Dawson *et al.* 1994; New England Biolabs) and both proteins were buffered at pH 7. A precipitate forms at around 0.1 mM of *bs*PcrA so the ligation mixture was left at just below 0.1 mM of *bs*PcrA overnight at 4 °C or at room temperature. Two faint bands, A and B, are seen by SDS-PAGE running at the same molecular weight as wild type *bs*PcrA (see Figure 3.11). Band A was identified as *bs*PcrA by peptide mass mapping. However, the ligation reaction is inefficient and produces very small amounts of full-length *bs*PcrA.

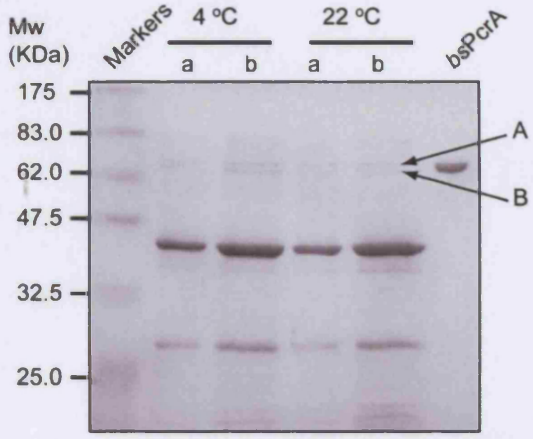

Figure 3.11. Ligation of *bs*PcrA1 and *bs*PcrA2 to form full length *bs*PcrA

The *bs*PcrA1 and *bs*PcrA2 proteins were concentrated together and incubated overnight at 4°C or at room temperature.

a= 5ul reaction mixture loaded onto gel, b= 10ul reaction mixture loaded onto gel.

3.2.8 Other *bs*PcrA N-terminal intein fusion proteins

Only small amounts of soluble *bs*PcrA1-*Mxe*GyrAi fusion protein could be made and the intein cleavage was inefficient. Only a very small amount of full length *bs*PcrA could be formed from ligation. Different N-terminal fragments of *bs*PcrA were fused to an intein in the hope that one of them would be soluble and the intein cleavage would, therefore, be more efficient. The crystal structure of *bs*PcrA was used to help in selecting fragments that did not end in the middle of any secondary structure (see Figure 3.12).

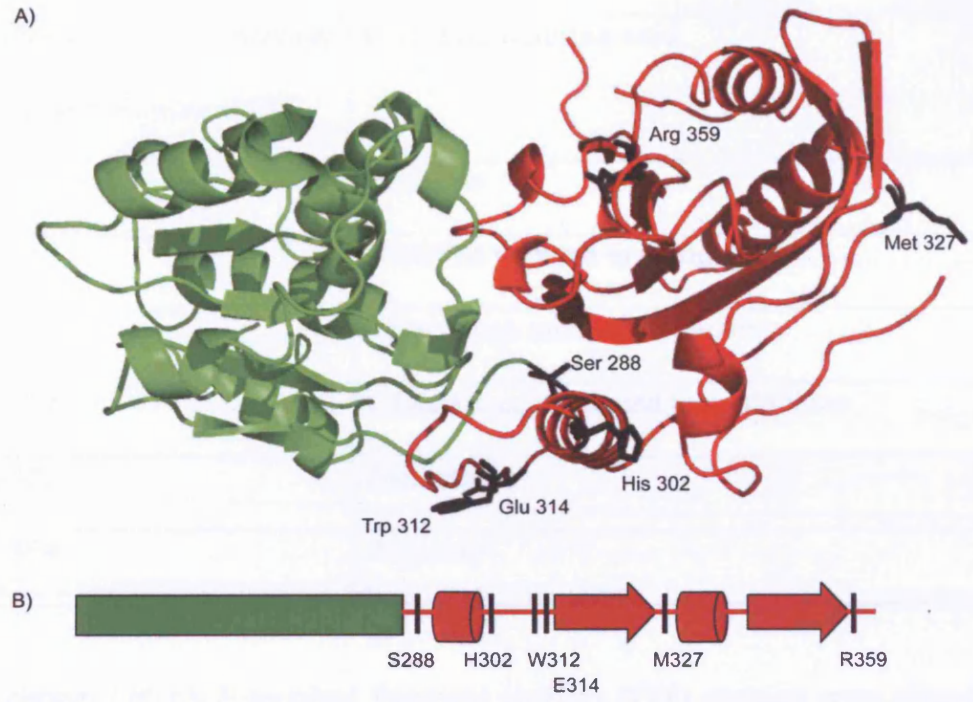

Figure 3.12. Potential intein fusion points within *bs*PcrA

Only domains 1A and 2A are shown for clarity. Residues labelled in black were chosen to fuse the N-terminal *bs*PcrA fragment to *Mxe*GryAi.

The potential C-terminal amino acid of the fragment was also considered as the intein cleavage efficiency is affected by the amino acid the intein is fused to. An appropriate mutation was made if the fragment did not end in a suitable amino acid and the mutation was reasonable in terms of how similar the amino acids were. None of the amino acids at potential fusion points were suitable for the *Mth*RIR1i (New England Biolabs), neither was it reasonable to mutate any of the amino acids as this would result in a large change to the residue. The fragments chosen for fusion with *Mxe*GyrAi are shown in Table 3.2 below.

Table 3.2. The *bs*PcrA N-terminal fragments chosen to be fused via C-terminal to *Mxe*GryAi

***bs*PcrA N-terminal fragment (amino acids)**	**C-terminal amino acid**
1-302	Histidine
1-302	Histidine mutated to methionine
1-312	Tryptophan mutated to tyrosine
1-314	Glutamic acid mutated to methionine
1-327	Methionine
1-359	Arginine

The chosen *bs*PcrA N-terminal fragment (*bs*PcrA NTF) proteins were cloned into pTWIN1. Table 3.3 below shows the expression trial conditions for the *bs*PcrA NTF–*Mxe*GyrAi fusion proteins. All apart from 1-E314M-*Mxe*GyrAi expressed, but none were soluble.

Table 3.3. Expression trials for *bs*PcrA N-terminal fragment *Mxe*GyrAi fusion proteins

***E.coli* Strain**	**Temperature (°C) and length of induction (hrs)**	**Concentration of IPTG (mM)**
BL21 2+	3 hrs, 37°C	1
	12-18 hrs 18°C	1
		0.2
Tuner 2+	3 hrs, 37°C	1
	12-18 hrs 18°C	1
		0.2

3.3 An attempt to make an *Archaeoglobus fulgidus* PCNA dimer

3.3.1 Introduction

Proliferating cell nuclear antigen (PCNA) is a processivity factor for DNA polymerase (DNA pol). The PCNA acts as a sliding clamp on DNA to keep DNA pol tethered to the DNA. PCNA is loaded onto DNA by the Replication Factor C (RFC) complex. The crystal structures of *Pyrococcus furiosus* and human PCNA revealed that three PCNA monomers form a ring structure (Bruck and O'Donnell 2001; Gulbis *et al.* 1996; Matsumiya *et al.* 2001). A trimeric ring model has also been proposed based on the crystal structure of the *Archaeoglobus fulgidus* PCNA monomer (Chapados *et al.* 2004) (see Figure 3.13). Each monomer contains two structurally homologous domains joined by a flexible linker resulting in a trimeric ring with pseudo-six-fold symmetry (see Figure 3.13).

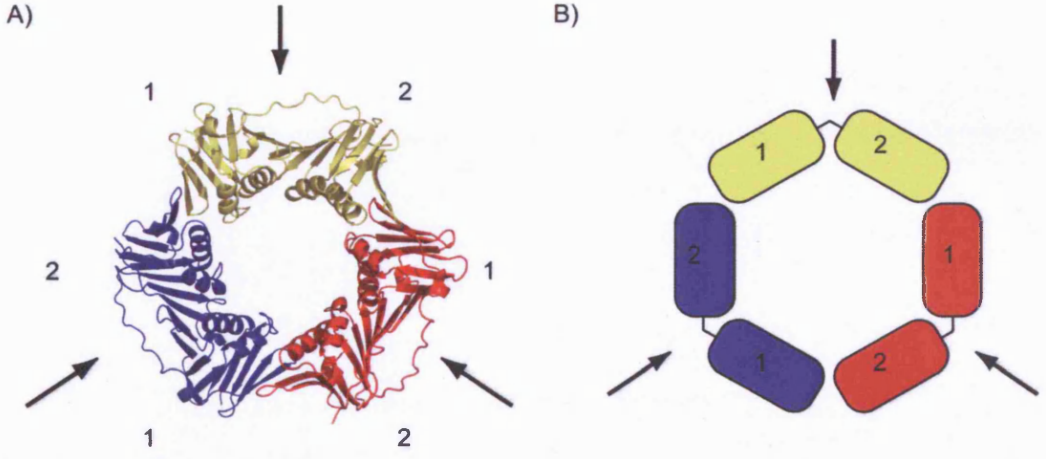

Figure 3.13. The crystal structure of *af*PCNA

The ring is made of an *af*PCNA homotrimer. The arrows point to flexible linkers that join the two homologous domains, 1 and 2, of one PCNA monomer together.

The aim of this part of the thesis was to address the question of how PCNA is loaded onto DNA. There are several ways that PCNA could be loaded to form a trimeric ring around DNA. Two main theories are the assembly theory and the ring break theory. The assembly theory is based on PCNA monomers, or a PCNA monomer and dimer, being assembled by RFC into a PCNA trimeric ring on DNA. PCNA may also self assemble into a trimeric ring without the aid of RFC; this may happen with the trimeric ring assembling directly onto the DNA. The ring break theory suggests that RFC may take the pre-made trimer, break one or more interfaces, and then load the PCNA onto the DNA forming the final trimer configuration (see Figure 3.14).

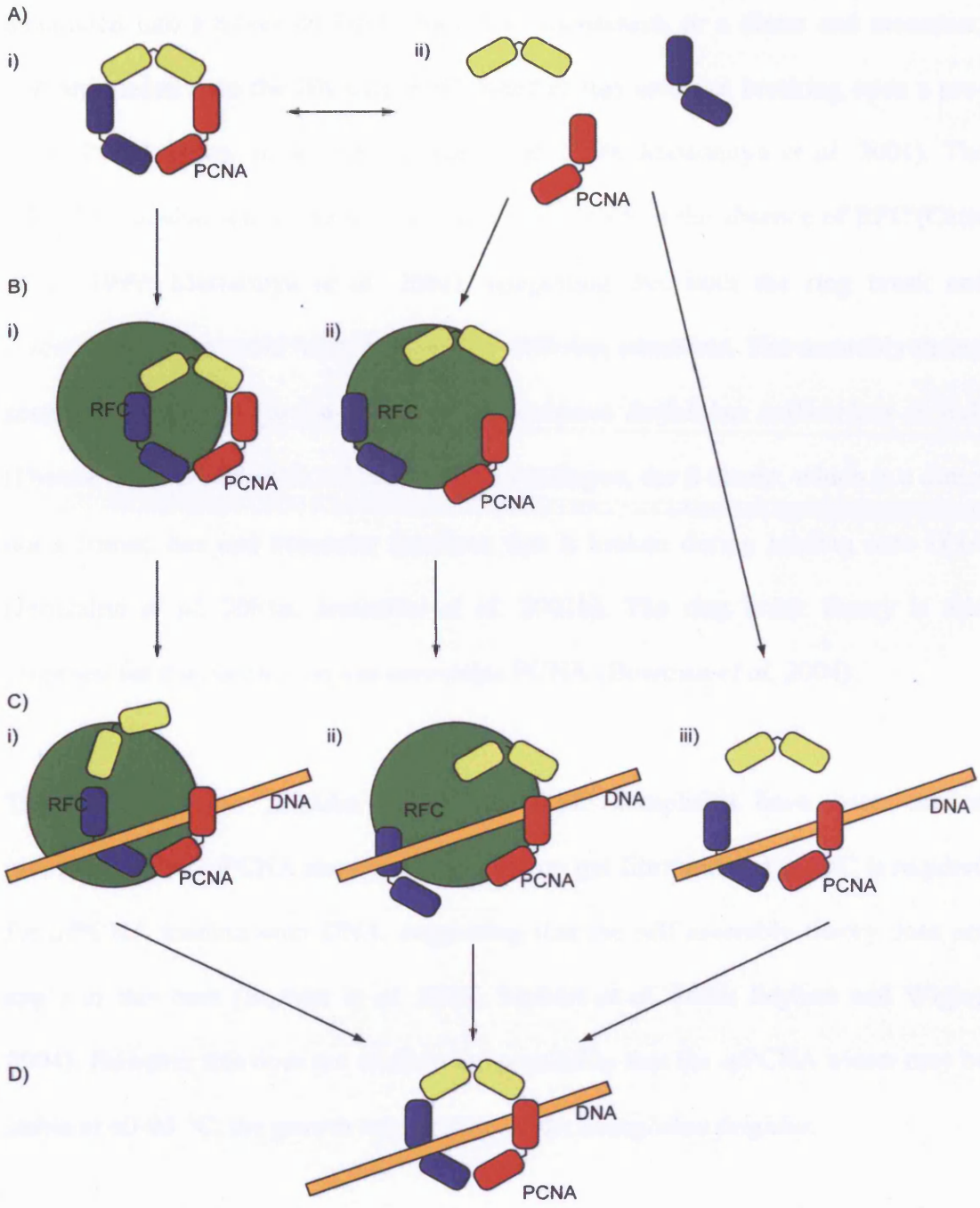

Figure 3.14. Possible ways for a PCNA trimer to be loaded onto DNA

A) Possible PCNA states, i) homotrimer, ii) monomers or dimer and monomer.

B) RFC may bind, i) a pre-formed PCNA trimer, ii) PCNA monomers, or dimer and monomer.

C) PCNA loading onto DNA by i) a RFC in a ring break mechanism, ii) RFC in a trimer assembly mechanism, iii) a PCNA trimer self assembly mechanism.

D) The PCNA homotrimer on DNA.

The PCNA from the archaeon *Pyrococcus furiosus* (*pf*PCNA) appears to be assembled into a trimer on DNA from three monomers, or a dimer and monomer, that are loaded onto the DNA by RFC; whether this involves breaking open a pre-made PCNA trimer is not clear (Cann *et al.* 1999; Matsumiya *et al.* 2001). The *pf*PCNA can also self assemble into a trimer on DNA in the absence of RFC (Cann *et al.* 1999; Matsumiya *et al.* 2001), suggesting that both the ring break and assembly theories could happen *in vivo* in differing situations. The assembly theory seems to be the case in the PCNA in the archaeon *Sulfolobus solfataricus* as well (Dionne *et al.* 2003). In *E.coli* the PCNA homologue, the β clamp, which is a dimer not a trimer, has one monomer interface that is broken during loading onto DNA (Jeruzalmi *et al.* 2001a; Jeruzalmi *et al.* 2001b). The ring break theory is also proposed for the *Saccharomyces cerevisiae* PCNA (Bowman *et al.* 2004).

The *Archaeoglobus fulgidus* PCNA and RFC complexes have been studied previously. The *af*PCNA runs as a monomer on gel filtration and *af*RFC is required for *af*PCNA loading onto DNA, suggesting that the self assembly theory does not apply in this case (Seybert *et al.* 2002; Seybert *et al.* 2006; Seybert and Wigley 2004). However this does not exclude the possibility that the *af*PCNA trimer may be stable at 60-95 °C, the growth temperature of *Archaeoglobus fulgidus.*

The specific aim of this project was to create an *af*PCNA dimer in order to investigate the loading properties using biochemical assays, and possibly for use in trying to crystallise an *af*PCNA/RFC complex. A covalently linked *af*PCNA dimer would allow investigations into whether three PCNA monomers are required by RFC in order to load a PCNA trimer on DNA, or whether RFC can load the linked dimer and monomer. This would provide information as to whether the RFC breaks

open a PCNA trimer, or assembles three PCNA monomers, in order to load trimeric PCNA onto DNA.

Steric strains could be introduced by directly fusing two PCNA monomers together that could alter or prevent the final trimeric ring conformation (see Figure 3.15). To try and avoid this a flexible linker could be introduced in between the two monomers to allow one monomer to move relative to the other.

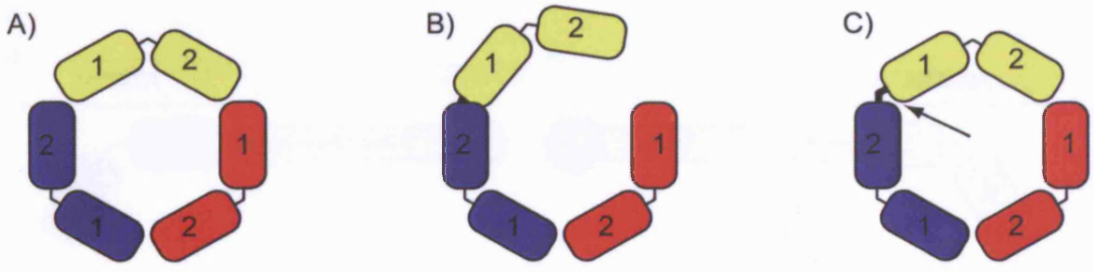

Figure 3.15. Fusing two PCNA monomers

A) A schematic of the PCNA trimer ring. Each PCNA monomer is coloured differently.
B) Directly fusing two PCNA monomers may alter or prevent the final trimer ring conformation.
C) Adding a flexible linker should allow relative movement between the fused PCNA monomers. The arrow indicates the added linker.

PCNA itself provided an amino acid sequence for such a linker, as a PCNA monomer has two homologous domains joined by a flexible linker. The *af*PCNA linker was not used because of possible recombination events between the introduced linker and the original during cloning; instead the flexible linker of *pf*PCNA was used. Based on the crystal structure of *pf*PCNA the flexible linker was defined as amino acids 116-129 and could be introduced by PCR onto an *af*PCNA monomer. The overall structure of the dimer would, therefore, resemble four homologous sub-domains all linked in a similar way, and should not produce any steric strains on the structure of the closed ring.

An attempt was made to clone two *af*PCNA genes next to one another, but this failed to work, probably due to recombination events within the cell between the two identical genes placed next to one another, despite the use of a $RecA^{mut}$ *E.coli* cloning strain. The IMPACT-TWIN system provided another way to create the dimer using intein-mediated protein ligation of two *af*PCNA monomers. (see Figure 3.16). It was decided to add the *pf*PCNA linker to the N-terminus of the *af*PCNA fused to *Ssp*DnaBi so that the cysteine required for IPL could be introduced to the N-terminus of the *pf*PCNA linker rather than the N-terminus of the *af*PCNA.

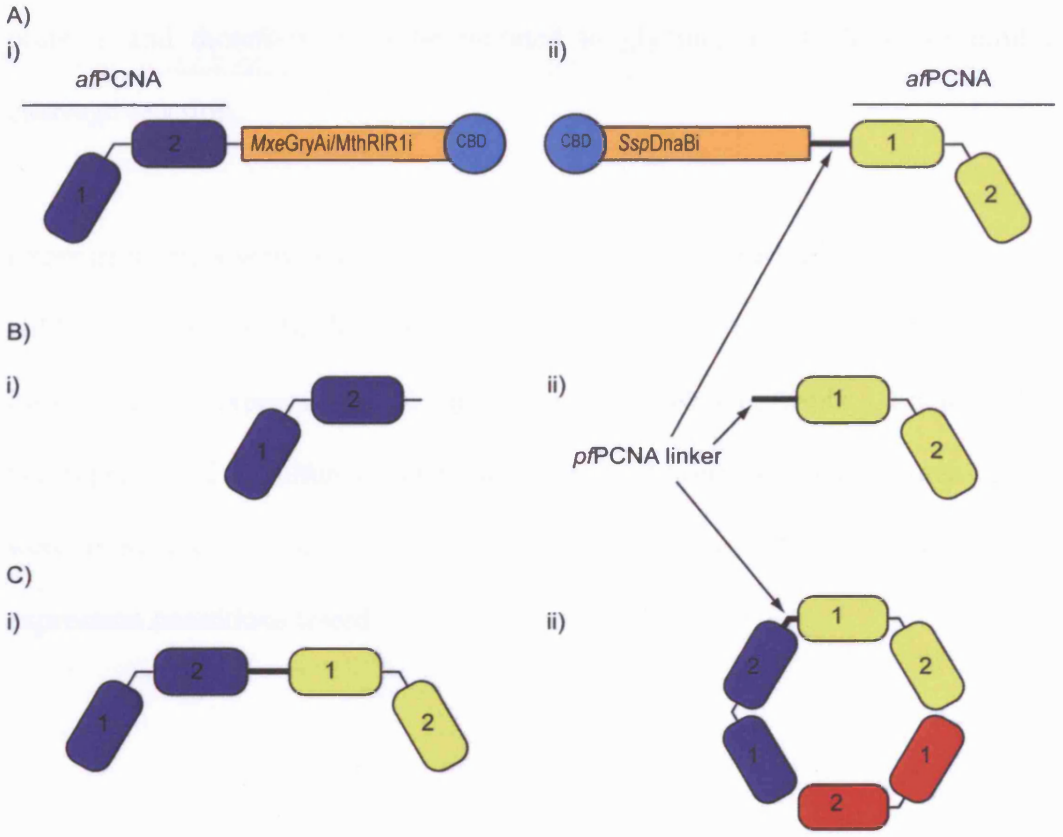

Figure 3.16. Strategy for creating an *af*PCNA dimer

Each *af*PCNA monomer is coloured differently.

A) *af*PCNA-intein fusion proteins to be made. i) The first *af*PCNA monomer, comprising two homologous domains, 1 and 2, joined by a flexible linker is fused via its C-terminus to *Mxe*GryAi (or *Mth*RIR1i). ii) The second *af*PCNA monomer with the *pf*PCNA linker is fused via its N-terminus to *Ssp*DnaBi.

B) Protein products after intein cleavage. i) *af*PCNA with a C-terminal thiol group. ii) *af*PCNA fused with the N terminal *pf*PCNA linker that has an N-terminal cysteine.

C) i) The final ligation product of two *af*PCNA monomers joined by a flexible linker from *pf*PCNA. ii)The *af*PCNA trimer comprising of the ligated dimer and another monomer.

3.3.2 Cloning and expression of *af*PCNA-*Mxe*GyrAi and *af*PCNA-*Mth*RIR1i

The *af*PCNA gene was cloned into pTWIN1 and pTWIN2 to produce constructs with *af*PCNA fused directly to the *Mxe*GyrAi or *Mth*RIR1i respectively. Previous work has shown that a C-terminal glutamate inhibits cleavage from *Mxe*GyrAi and *Mth*RIR1i fusions (New England Biolabs). Sequence comparisons show that this glutamate residue, which is present in *af*PCNA, is not conserved in other PCNA proteins and therefore could be mutated to glycine, which does not inhibit the cleavage reaction.

Expression trials were undertaken. The B834 2+ *E.coli* cell line was used to express wild type *af*PCNA (Seybert *et al.* 2002) and was used for the fusion protein. The fusion proteins expressed in the 5ml expression trials (see Table 3.4 below), but did not express in 1 L cultures under the same conditions. Smaller cultures of 200 ml were tried and the fusion proteins expressed. Table 3.4 below summarises the expression conditions tested.

Table 3.4. Summary of expression trials of *af*PCNA-*Mxe*GyrAi and *af*PCNA-*Mth*RIR1i

***E.coli* Strain**	**Induction conditions**	**Culture volume (ml)**	***af*PCNA-*Mxe*GyrAi**	***af*PCNA-*Mth*RIR1i**
			Expression and solubility	**Expression and solubility**
BL21 +	1mM IPTG, 3hrs, 37°C	5	-/+	-/+
	1mM IPTG, 12-18hrs, 18°C	5	-/+	-/+
BL21 3+	1mM IPTG, 3hrs, 37°C	5	-/+	-/+
	1mM IPTG, 12-18hrs, 18°C	5	-/+	-/+
B834 2+	1mM IPTG, 3hrs, 37°C	5	+	+
		200	+	+
		1000	-	-
	1mM IPTG, 12-18hrs, 18°C	5	+	+
		200	+	+
		1000	-	-

3.3.3 Purification of *af*PCNA from *af*PCNA-*Mxe*GyrAi and *af*PCNA-*Mth*RIR1i

Small-scale purification tests were tried, as described earlier in section 3.2.3, to see if *af*PCNA could be purified from either the *af*PCNA-*Mxe*GryAi or *af*PCNA-*Mth*RIR1i fusion proteins. The *af*PCNA-*Mth*RIR1i fusion protein cleaves well in the small-scale tests, as expected with glycine at the fusion point, so *af*PCNA was

purified from *af*PCNA-*Mth*RIR1i using a chitin column (see Figure 3.17 A). The *af*PCNA-*Mxe*GyrAi fusion protein does not cleave as well, but *af*PCNA can also be purified from this fusion protein (see Figure 3.17 B).

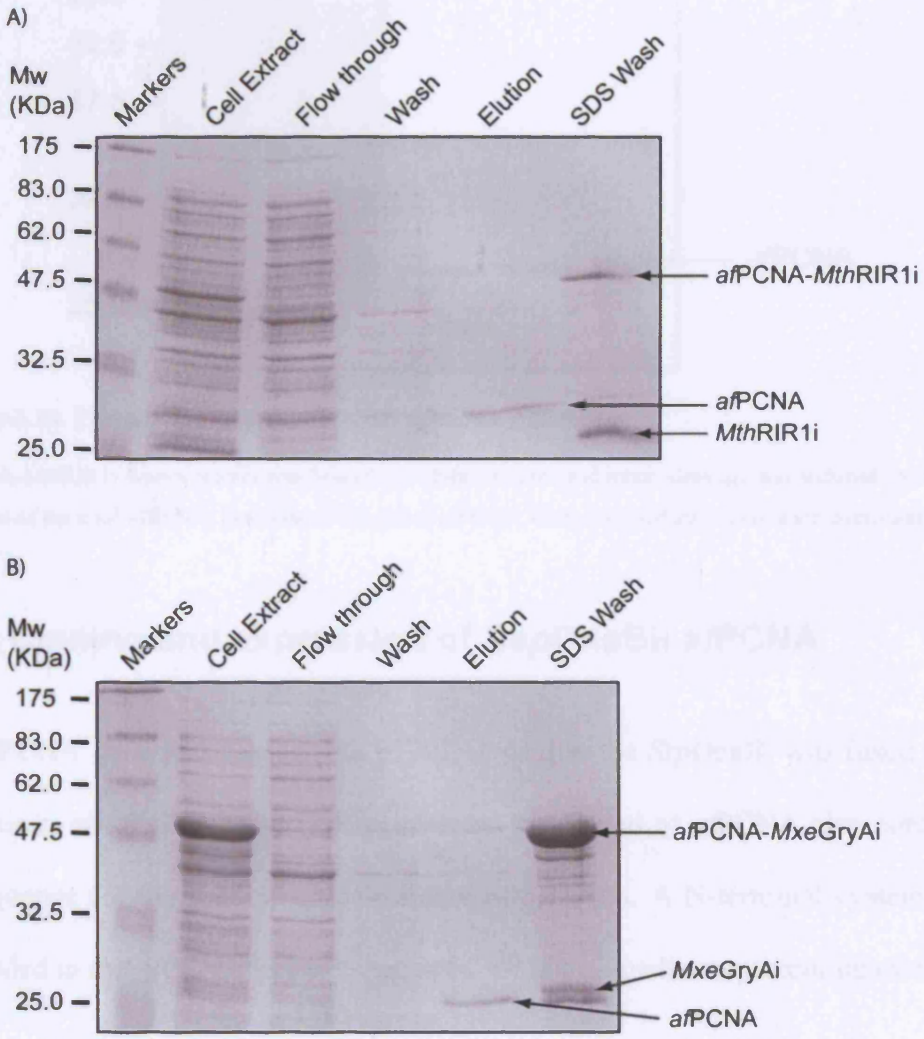

Figure 3.17. Small-scale purification tests of *af*PCNA fused to either *Mxe*GryAi or *Mth*RIR1i

A) *af*PCNA-*Mth*RIR1i. B) *af*PCNA-*Mxe*GryAi.

The *af*PCNA-*Mth*RIR1i fusion protein was chosen for a large-scale purification (see Figure 3.18) because *Mth*RIR1i cleaved better than *Mxe*GyrAi .

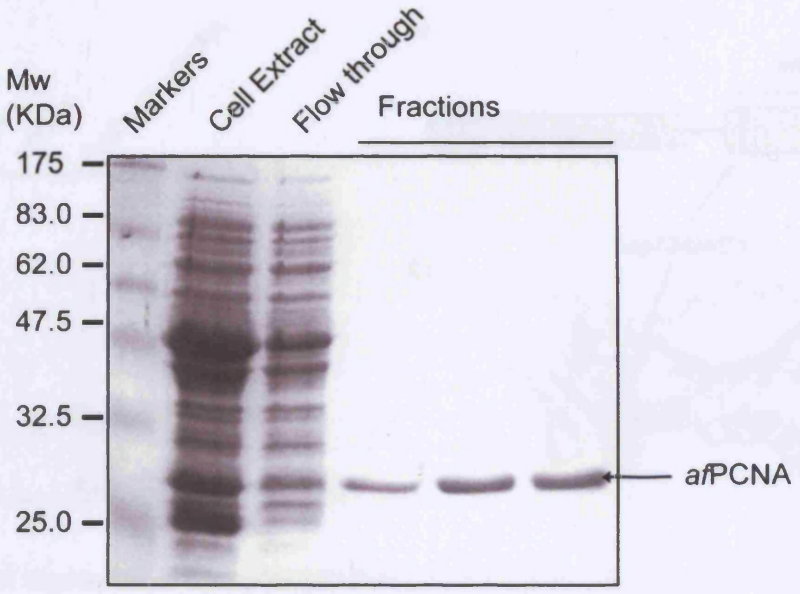

Figure 3.18. Purification of *af*PCNA from *af*PCNA-*Mth*RIR1i

*af*PCNA-*Mth*RIR1i fusion protein was bound to a chitin column and intein cleavage was induced by the addition of the thiol MESNA. Fractions of the eluted *af*PCNA were collected after incubation overnight.

3.3.4 Cloning and expression of *Ssp*DnaBi- *af*PCNA

The *af*PCNA gene was cloned into pTWIN1 so that the *Ssp*DnaBi was fused to the N-terminus of *af*PCNA. The PCR primer for the 5' end of *af*PCNA also contained the sequence for the flexible domain linker of *pf*PCNA. A N-terminal cysteine was also added to the *pf*PCNA linker as required for intein-mediated protein ligation.

Two expression products can be seen by SDS-PAGE (see Figure 3.19). Peptide mass mapping identified the smaller band as a fragment of *af*PCNA and the larger band as the intein (see Figure 3.19).

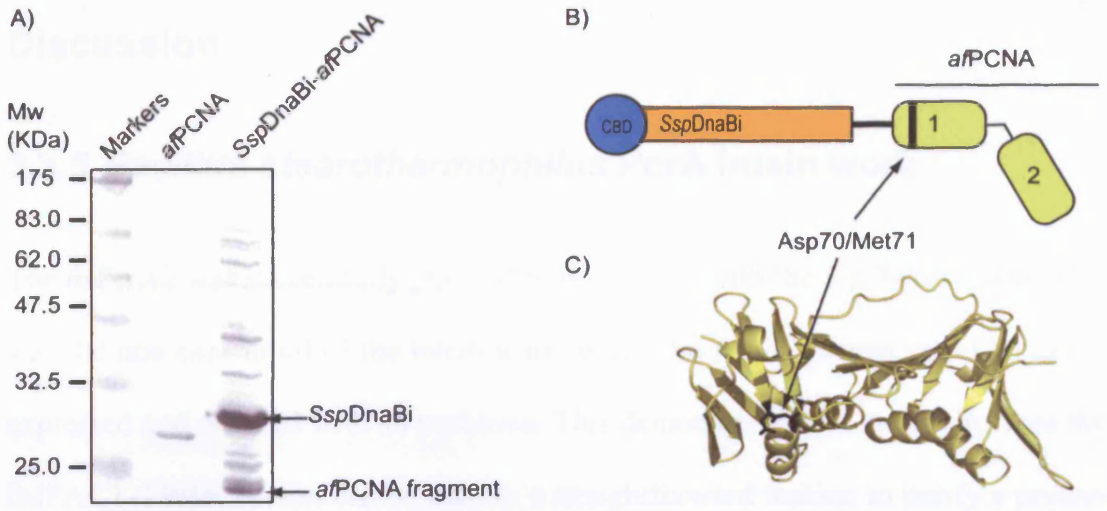

Figure 3.19. Expression of *Ssp*DnaBi-*af*PCNA

A) Expression of *Ssp*DnaBi-*af*PCNA compared to purified *af*PCNA.

B) A schematic of the fusion protein with residues at which proteolysis is occurring labelled in black.

C) Crystal Structure of *af*PCNA with the residues at which proteolysis is occurring labelled in black.

Various expression conditions were tested (see Table 3.5), but all gave the same expression pattern. The N-terminus of the *af*PCNA was sequenced and shows that the cleavage is taking place after aspartate 70 in the sequence TIGVD/MDRIF of the *af*PCNA (see Figure 3.19).

Table 3.5 Expression trial conditions of *Ssp*DnaBi-*af*PCNA.

E.coli Strain	Temperature and length of induction
B834 2+	3hrs, 37°C
	12-18hrs, 18°C
BL21 2+	3hrs, 37°C
	12-18hrs, 18°C
BL21 +	3hrs, 37°C
	12-18hrs, 18°C
BL21 3+	3hrs, 37°C
	12-18hrs, 18°C

All conditions used 1 mM IPTG

3.3.5 *Bacillus stearothermophilus* PcrA intein work

The *bs*PcrA2 was successfully purified from a fusion with the *Ssp*DnaBi intein. This was the one case in all of the intein work where the fusion protein could be easily expressed and purified with no problems. This demonstrates that in certain cases the IMPACT-TWIN system can be used in a straightforward fashion to purify a protein using a self-cleavable affinity tag.

The lack of solubility of *bs*PcrA1-*Mxe*GyrAi and *bs*PcrA1-*Mth*RIR1i fusion proteins was a problem. There is no obvious reason why *bs*PcrA1 should be insoluble other than the high levels of expression, so it was decided to try and reduce the amount of expression using Tuner cells. Tuner cells have a mutated lac permease (lacY) protein this means IPTG uptake is uniform in the cell culture and the amount of expression is dependant on the concentration of IPTG. Tuner cells can therefore be induced to express high or low levels of protein by varying the concentration of IPTG used. This idea worked well and a small amount of protein was soluble when expressed in the Tuner *E.coli* cells using 0.2 mM IPTG.

Once soluble *bs*PcrA1-*Mxe*GyrAi protein had been made little or no intein cleavage was observed this could not have been predicted, as there were no reports suggesting that a serine at the cleavage site would inhibit the cleavage reaction. Other studies have shown that having a methionine at the cleavage site lead to high levels of intein cleavage and muating serine 288 to methionine (S288M) did indeed increase the amount of cleavage. There is not a lot of work defining which amino acids at the

cleavage point are better then others for allowing the cleavage reaction to work. The information available comes from individual reports of different proteins being cleaved from the inteins. It is reasonable to speculate that the protein fold and sequence context of the amino acid will play an important part in determining the cleavage efficiency. While some conclusions can be drawn as to which amino acids are best for the intein cleavage it seems that trial and error is the best approach when working with a protein that has not been used with inteins before.

Some *bs*PcrA1 became insoluble when cleaved from *Mxe*GryAi reducing the yield. This suggests that the insolubility of the *bs*PcrA1-*Mxe*GryAi fusion protein was due to the *bs*PcrA1 being intrinsically insoluble rather than just the high levels of protein expression.

The ligation reaction between the *bs*PcrA1 and *bs*PcrA2 was not very successful or reproducible. This meant that only a very small amount of full-length *bs*PcrA was created in amongst lots of contaminants. The *bs*PcrA halves could only be concentrated to around 0.1 mM before a precipitate formed. This may not be high enough for the ligation reaction to proceed. The precipitation is likely to be the *bs*PcrA1 forming insoluble aggregates as *bs*PcrA1 seemed intrinsically insoluble where as *bs*PcrA2 did not; if *bs*PcrA1 was forming aggregates then this would prevent *bs*PcrA2 associating with *bs*PcrA1 to allow the ligation reaction to occur.

During this time the RecD work (described in Chapters 4 and 5) had started to yield results and it was decided that the RecD project should be given priority because the *bs*PcrA work had, to this date, not produced any firm data relating to the original question of what *bs*PcrA domains 1B and 2B contribute to duplex DNA unwinding.

If time had permitted the *bs*PcrA work would have been revisited to try and purify full length *bs*PcrA ligated from *bs*PcrA1 and *bs*PcrA2.

Ligation of *bs*PcrA1 and *bs*PcrA2 had not produced a positive control and it was decided not to try and make the domain deletion mutants. In retrospect, a *bs*PcrA mutant that contained the cysteine that would have been present from purification from the *Ssp*DnaBi- *bs*PcrA fusion protein could have been made using site directed mutagenesis of the wild-type *bs*PcrA gene and used as a pseudo positive control. How ever this would not have been ideal and any changes in activity of the domain deleted *bs*PcrA could still be argued to have been a result of the method of making them.

3.3.6 *Archaeoglobus fulgidus* PCNA intein work

This part of the intein work provided a different set of problems to the *bs*PcrA work. Firstly, in trying to express both the *af*PCNA fusion proteins and secondly it was the *Ssp*DnaBi-*af*PCNA fusion protein that could not be made, where as in the *bs*PcrA work *bs*PcrA2 was successfully purified from the *Ssp*DnaBi-*bs*PcrA2 fusion protein. In the case of *af*PCNA-*Mxe*GyrAi/*Mth*RIR1i the fusion proteins expressed in 5 ml and 200 ml cultures, but not in 1 L cultures. There does not seem to be any good reason for this, the aeration of the culture may have some impact but it is hard to see how a 1 L culture could be so deficient in aeration as to stop expression of the protein especially as the 1 L culture grew at the same rate as the 200 ml culture.

As a tool for purifying *af*PCNA the intein system works very well. The *af*PCNA-*Mth*RIR1i fusion protein allowed the purification of *af*PCNA from a single column using the chitin-binding domain as an affinity tag instead of three columns as

previously described (Seybert *et al.* 2002). The power of this method is that no tag is left attached to the protein. Using another reducing agent other than MESNA, which is required for the ligation reaction, could increase the cleavage efficiency of the intein. If the IMPACT-TWIN system was just being used to purify *af*PCNA other reducing agents could be tried to see if the yield of *af*PCNA could be increased.

There was no attempt to ligate two *af*PCNA monomers together. This is due to the fact that the *af*PCNA monomer with an N-terminal cysteine and *pf*PCNA linker could not be purified, because the full-length *Ssp*DnaBi - *af*PCNA fusion protein could not be expressed. The expression led to two products as described earlier. One product was a fragment of *af*PCNA (amino acids 71-245) and the second was the *Ssp*DnaBi intein, presumably fused to the first 70 amino acids of *af*PCNA, although this wasn't detected in the peptide mass mapping. This proteolysis is unlikely to be caused by the *Ssp*DnaBi as it is reported to cleave best at cysteine, glycine and serine amino acids (New England Biolabs); neither has the intein been reported to cleave anywhere else other then the fusion site between the intein and the target protein. The proteolysis occurred despite the use of the protease inhibitors PMSF, pepstatin, leupeptin and EDTA that inhibit the serine, acid, thiol and mettalo proteases respectively. The site of assumed proteolysis could have been mutated to try and stop the problem. This was not attempted due to time constraints and mutation might disturb the secondary structure of the protein especially in this case where proteolysis is occurring in between a β sheet and an α helix.

In conclusion this work failed to create *bs*PcrA domain deletion mutants or a *af*PCNA dimer. However, the project did demonstrate that the IMPACT-TWIN system is capable of purifying proteins using a self-cleavable affinity tag.

4 Biochemical analysis of *Deinococcus radiodurans* RecD2

4.1 Introduction

4.1.1 Summary

This part of the thesis describes the identification, purification and biochemical and mutational analysis of *Deinococcus radiodurans* RecD2 (*dr*RecD2) in order to gain insights into the molecular mechanism of Superfamily 1Bα (SF1Bα) helicase activity. The focus of this study was the molecular mechanism of a Superfamily 1Bα helicase and not the *in vivo* function of *dr*RecD2. The main conclusions are, firstly, that a structurally conserved loop in the RecD2 family could be acting as a molecular pin for splitting open the duplex DNA ahead of the motor domains. Secondly, that the N-terminal domain of *dr*RecD2 may be acting as a regulatory domain that inhibits helicase activity. Lastly, that a conserved motif (Motif A) may not be involved in the helicase activity of *dr*RecD2.

4.1.2 The Superfamily 1Bα RecD helicase

There are two types of RecD helicases as described in the introduction to this thesis (see Chapter 1, section 1.10.3). The RecD enzymes that exist in organisms that have the RecB and RecC genes are classified as type 1 RecD (RecD1) helicases and are likely to be part of the RecBCD1 complex as a *bona fide* RecD. The RecD enzymes that exist in organisms that do not possess RecB or RecC genes are classified as type 2 RecD (RecD2) helicases. A RecD2 helicase, therefore, cannot be part of a

RecBCD1 complex and these are not therefore *bona fide* RecD enzymes, but are homologous SF1Bα helicases.

A histidine tagged construct of *E.coli* RecD1 (*E.coli* RecD1^{his}) has been shown to have ssDNA-dependent ATP hydrolysis and 5' – 3' helicase activities *in vitro* in isolation from the RecBCD1 complex (Chen *et al.* 1997; Dillingham *et al.* 2003). The purification of *E.coli* RecD1^{his} is not straightforward, because the protein is insoluble when expressed in the absence of the *E.coli* RecC and RecB proteins. The *E.coli* RecD1^{his} protein was refolded from inclusion bodies and was not able to fully restore the wild-type Chi-specific cleavage activity of a reconstituted holoenzyme comprising *E.coli* RecBC and RecD1^{his} (Dillingham *et al.* 2003). This indicates that not all of the protein was correctly folded. Further (unpublished) experiments suggest that only 10% of the *E.coli* RecD1^{his} is active (personal communication, Dillingham, M.).

4.2 Identification and purification of *Deinococcus radiodurans* RecD2

4.2.1 *E.coli* RecD1

In order to investigate the possibility of using *E.coli* RecD1 as a target for crystallisation it was cloned, expressed and purified based on previous work by Dillingham *et al.* (Dillingham *et al.* 2003) (see Figure 4.1).

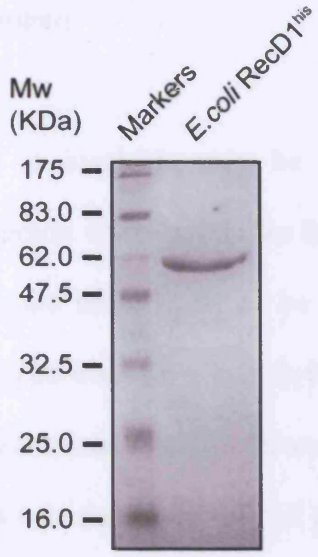

Figure 4.1. Purified *E.coli* RecD1^{his}

A 12% SDS-PAGE analysis showing the final purification product of *E.coli* RecD1^{his}

No ATP hydrolysis or helicase activities were observed from the protein despite several purification attempts (data not shown). Domain 1 of *E.coli* RecD1 interacts with RecC in the RecBCD complex, and it was reasoned that an *E.coli* RecD1 protein missing this domain might be soluble. Two N-terminal truncations of *E.coli* RecD1 missing domain one were cloned into expression vectors missing amino acids 1-109 and 1-126, *E.coli* RecD $^{\Delta 1\text{-}109}$ and *E.coli* RecD1$^{\Delta 1\text{-}126}$ in an attempt to make

soluble protein. Both truncated proteins were insoluble. It was decided to concentrate on identifying a novel RecD target for crystallisation.

4.2.2 Identification of *Deinococcus radiodurans* RecD2

In order to avoid the problems associated with refolding insoluble proteins it was decided to screen for a RecD helicase from another organism that, when cloned into an expression vector, would provide a high yield of soluble protein. A strategy was devised in order to maximise the chances of finding a RecD gene that would provide large quantities of soluble protein.

The reason *E.coli* RecD1 is insoluble may be because it has a very large hydrophobic surface that interacts with RecC in the RecBCD1 complex (Singleton *et al.* 2004). RecD2 helicases are more likely to be soluble than RecD1 helicases because RecD2 proteins have no need for a large hydrophobic surface interface area to interact with RecC. For this reason RecD2 helicases were prioritised in the screen. Another factor in the screen was the availability of genomic DNA, so some RecD1 helicases were also cloned. All the screened RecD helicases were cloned into expression vectors and underwent expression trials to determine if the clone would express soluble protein. The results of the screen are shown in Table 4.1.

Table 4.1. Screening for a soluble RecD helicase

Organism	Name	Size (KDa)	Clones made	Expression	Solubility
Deinococcus radiodurans	RecD2	76.4	Native	+	-
			C-terminal histidine tag	-/+	+
Streptococcus pneumoniae	RecD2	88.7	Native	+	-
Bacillus cereus	RecD2	87.4	Native	-	N/A
Lactobacillus plantarum	RecD2	93.0	Native	+	-
Desulfovibrio vulgaris	RecD2	82.6	Native	-	N/A
*Methanococcus jannaschii**	YF19_METAJ	138.6	Native	-	N/A
Heamophilus Influenzae	RecD1	72.8	Native	-	N/A
Mycobacterium smegmatis	RecD1	59.5	Native	+	-

* This is an archaeon and not a bacterium like the rest of the organisms in this list. The gene cloned from this organism is not a RecD gene, but it does possess a RecD-like domain.

The RecD clones underwent expression trials and either the proteins did not appear to express or the protein produced was insoluble. Initially it appeared that *Deinococcus radiodurans* RecD2 protein was expressed in the BL21 2+ cell line, but was insoluble (see Figure 4.2).

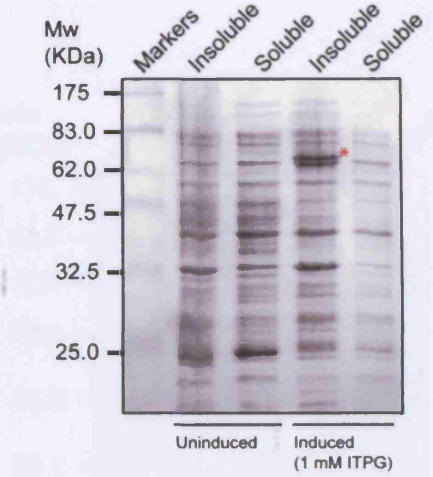

Figure 4.2. Initial expression trial of *Deinococcus radiodurans* RecD2

A 12% SDS-PAGE analysis showing the expression of insoluble *Deinococcus radiodurans* RecD2 indicated by *. The *E.coli* BL21 2+ with the *dr*RecD2 expression vector was grown to 0.6 $O.D_{600nm}$ and was induced with 1 mM IPTG for 3 hours. A control culture was not induced.

At the same time as this work was being performed data were published by Wang *et al.* that described a soluble C-terminal histidine tagged *Deinococcus radiodurans* RecD2 that had 5' – 3' helicase activity (Wang and Julin 2004). This study therefore cloned *Deinococcus radiodurans* RecD2 with a C-terminal histidine tag, (*dr*RecD2). Expression trials were performed, but no *dr*RecD2 protein was seen to be expressed (see Figure 4.3).

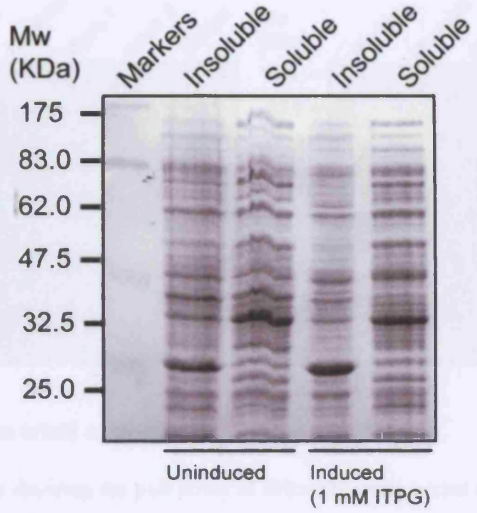

Figure 4.3. Expression trial of *dr*RecD2 with a C-terminal histidine tag

A 12% SDS-PAGE analysis showing the expression trial of C-terminally histidine tagged *Deinococcus radiodurans* RecD2. The *E.coli* BL21 2+ with the *dr*RecD2 expression vector was grown to 0.6 $O.D_{600nm}$ and was induced with 1 mM IPTG for 3 hours. A control culture was not induced.

It may have been the case that only very small amounts of *dr*RecD2 were expressed that could not be easily seen using SDS-PAGE analysis. In order to test for low levels of *dr*RecD2, small scale purification tests were carried using nickel agarose beads to pull down *dr*RecD2 from different cell lines. The *E.coli* BL21 Rosetta strain appeared to produce the best yield of *dr*RecD2 as analysed by SDS-PAGE (see Figure 4.4).

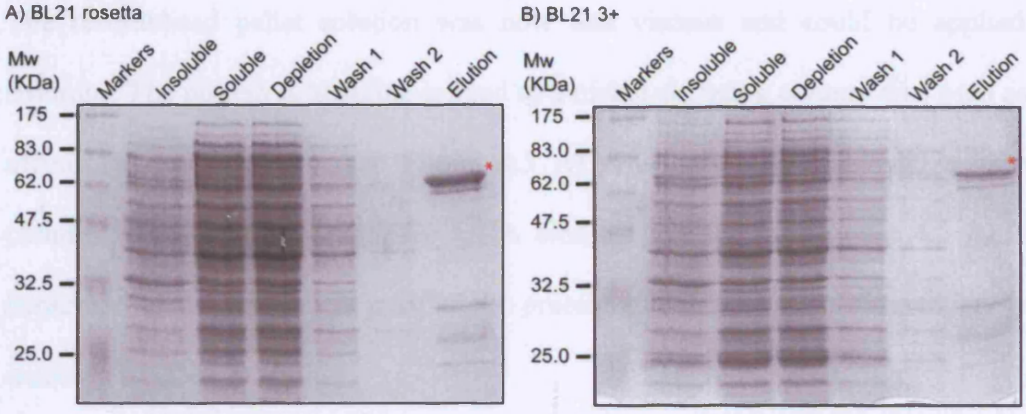

Figure 4.4. Purification trials of *dr*RecD2

12 % SDS-PAGE analysis showing the pull down of *dr*RecD2 using nickel agarose beads from A) BL21 Rosetta *E.coli* and B) BL21 3+ *E.coli*. The *E.coli* expression strains with the *dr*RecD2 expression vector were grown to 0.6 $O.D_{600nm}$ and were induced with 1 mM IPTG for 3 hours. The soluble fraction was isolated and the *dr*RecD2 was depleted from the fraction using nickel agarose beads. The beads were washed and then the *dr*RecD2 was eluted from the beads using imizadole as a competitor for nickel binding.

4.2.3 Purification of *Deinococcus radiodurans* RecD2

Based on the method of purification described by Wang *et al.* a high throughput and more stringent purification protocol was developed. Wang *et al.* utilised a nickel chelating column and a ssDNA cellulose column to purify *dr*RecD2. A ssDNA cellulose column is not suitable for a high throughput purification of large quantities of protein because the medium is single-use, so a heparin column was used instead. An extra purification step using a MonoQ anion exchange column was included to remove the last remaining contaminants to ensure the protein was sufficiently pure for crystallisation. An outline of the purification protocol is described here (see Materials and Methods for specific details).

The protein was over-expressed in ten 1 L cultures of *E.coli*. In order to reduce the viscosity of the resulting lysate the first step of purification was a 65% ammonium

sulphate precipitation that removes lipids, DNA and some protein from the lysate. The resolublised pellet solution was now less viscous and could be applied to columns. The protein is histidine-tagged so a nickel chelating column was used as an affinity purification step (see Figure 4.5 A). The next step employed a heparin column as an affinity column for DNA binding proteins (see Figure 4.5 B). The heparin column step further purifies the protein and also helps to remove any trace amounts of DNA nucleases.

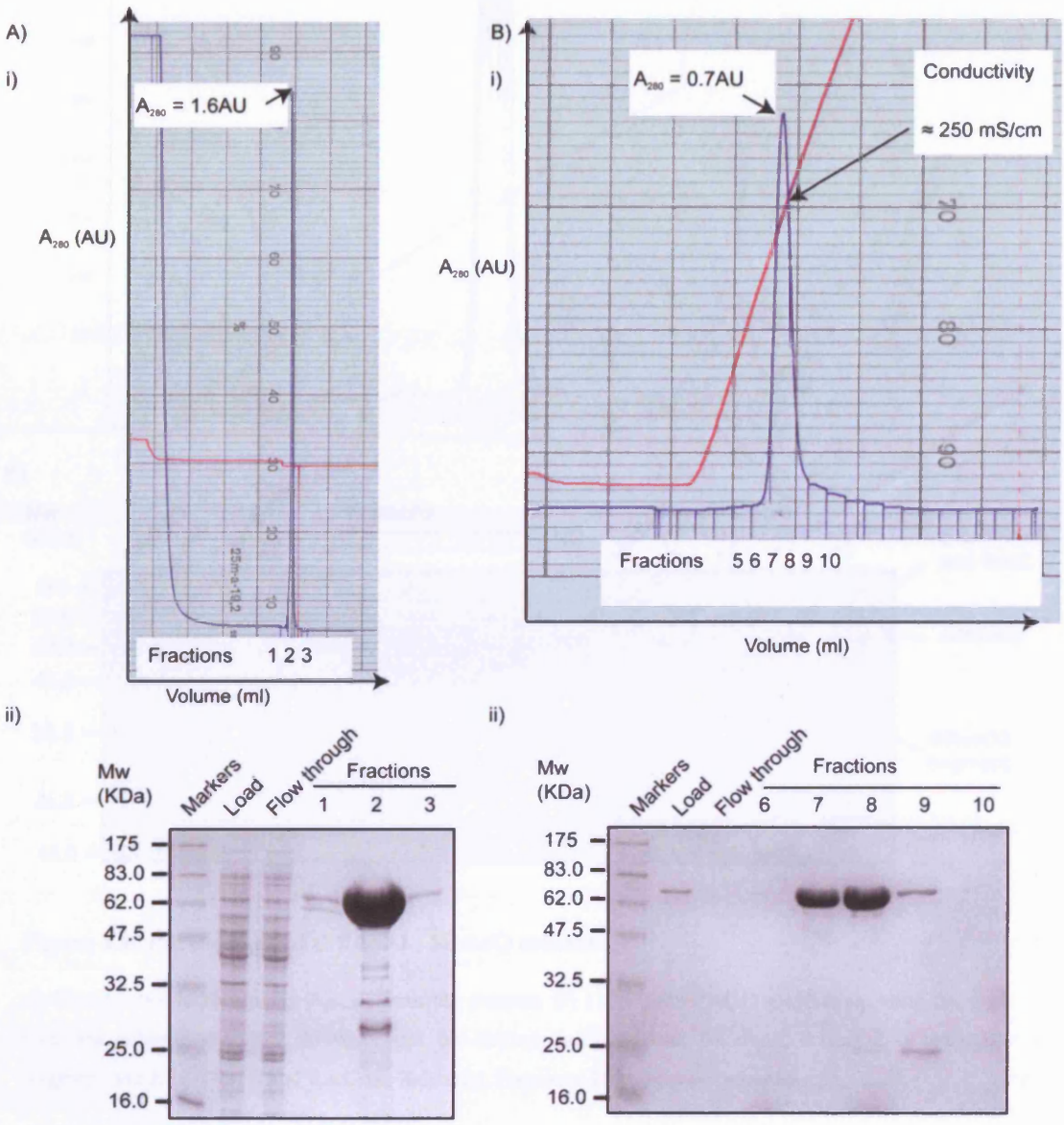

Figure 4.5. Purification of *dr*RecD2 using nickel chelating and heparin columns

A) Nickel chelating column. i) Elution profile measuring $A_{280\,nm}$ from the column. ii) A 12 % SDS-PAGE analysis showing the load onto the column, the flow through from the column and the peak fractions. Fraction 2 was applied to the heparin column.

B) Heparin column. i) Elution profile measuring $A_{280\,nm}$ from the column. ii) A 12 % SDS-PAGE analysis showing the load onto the column, the flowthrough from the column and the peak fractions. Fractions 7 and 8 were applied to the MonoQ column.

As there are still some contaminating proteins seen eluting with *dr*RecD2 from the heparin column a MonoQ column was used as an anion exchange purification step (see Figure 4.6).

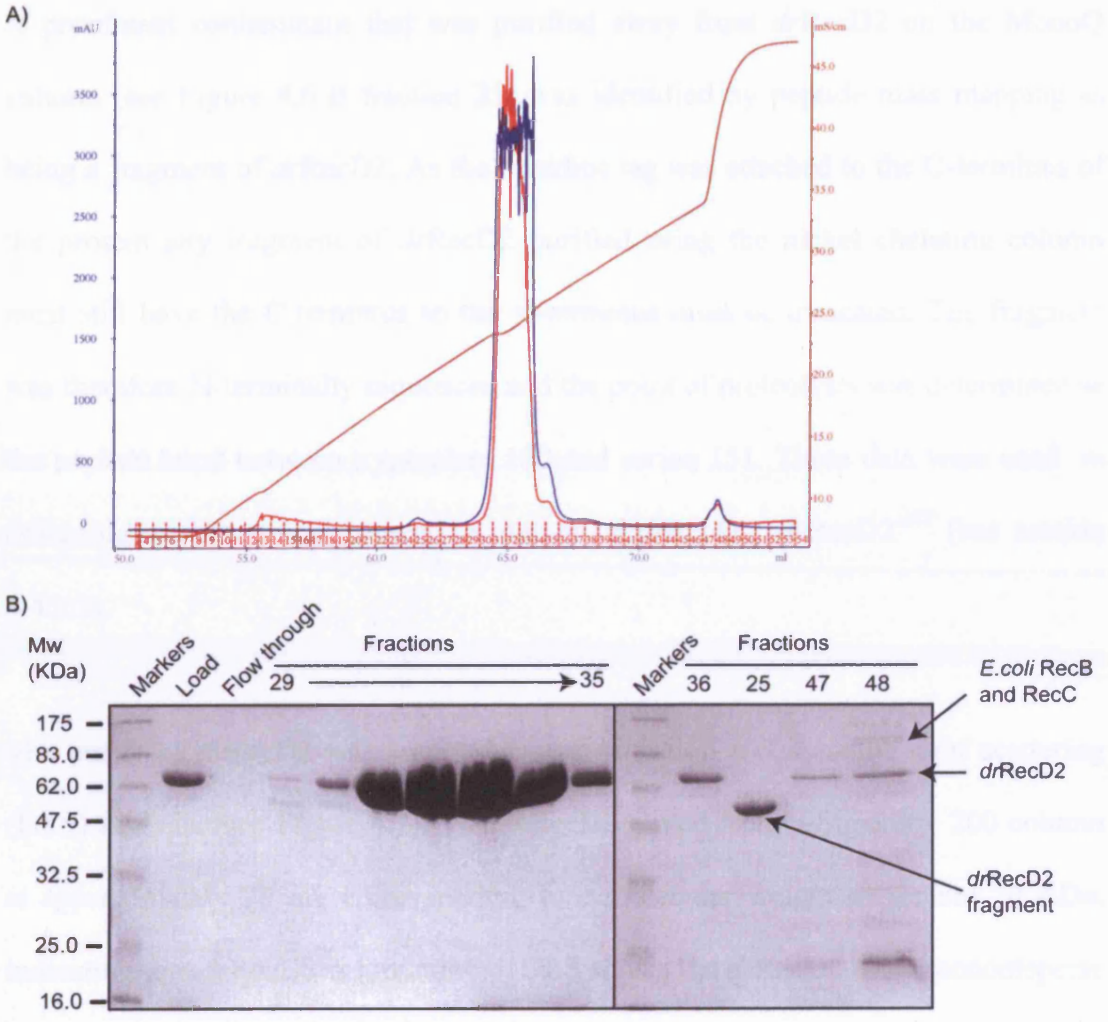

Figure 4.6. Purification of *dr*RecD2 , MonoQ column

A) Elution profile measuring $A_{280\ nm}$ from the column. B) 12 % SDS-PAGE analysis showing the load onto the column, the flow through from the column and the peak fractions, *dr*RecD2, a *dr*RecD2 fragment and *E.coli* RecB and RecC are indicated. Fractions 31 to 34 were selected.

A calculated pI of *dr*RecD2 was determined as 6.2 using the ProtParam tool of ExPASy (Gasteiger *et al.* 2003). The *dr*RecD2 containing fractions from the heparin column were loaded onto the MonoQ column at pH 7.5, giving a net negative surface charge and allowing it to bind to the MonoQ column at low sodium chloride concentrations. The MonoQ step also had the added benefit of concentrating the *dr*RecD2 as the protein is eluted as a sharp peak into a relatively small volume.

A prominent contaminant that was purified away from *dr*RecD2 on the MonoQ column (see Figure 4.6 B fraction 25) was identified by peptide mass mapping as being a fragment of *dr*RecD2. As the histidine tag was attached to the C-terminus of the protein any fragment of *dr*RecD2 purified using the nickel chelating column must still have the C-terminus so the N-terminus must be truncated. The fragment was therefore N-terminally sequenced and the point of proteolysis was determined as the peptide bond between tryptophan 150 and serine 151. These data were used in designing an N-terminal truncation mutant of *dr*RecD2, *dr*RecD2$^{\Delta NT}$ (see section 4.4.3.2).

The purity of *dr*RecD2 was assessed by gel filtration and dynamic light scattering (DLS) analysis (see Figure 4.7). The *dr*RecD2 eluted from a Superdex 200 column at approximately 70 ml, corresponding to a molecular weight of around 70 KDa, indicating that *dr*RecD2 is monomeric. DLS shows the *dr*RecD2 to be monodisperse as a single particle with a hydrodynamic radius of 7.86 nm, corresponding to an estimated molecular weight of 82 KDa, assuming a globular protein. These data suggest *dr*RecD2 is monomeric and does not form dimers or larger order aggregates in 20 mM Tris pH 7.5, 100 mM sodium chloride, 1 mM DTT and 1 mM EDTA.

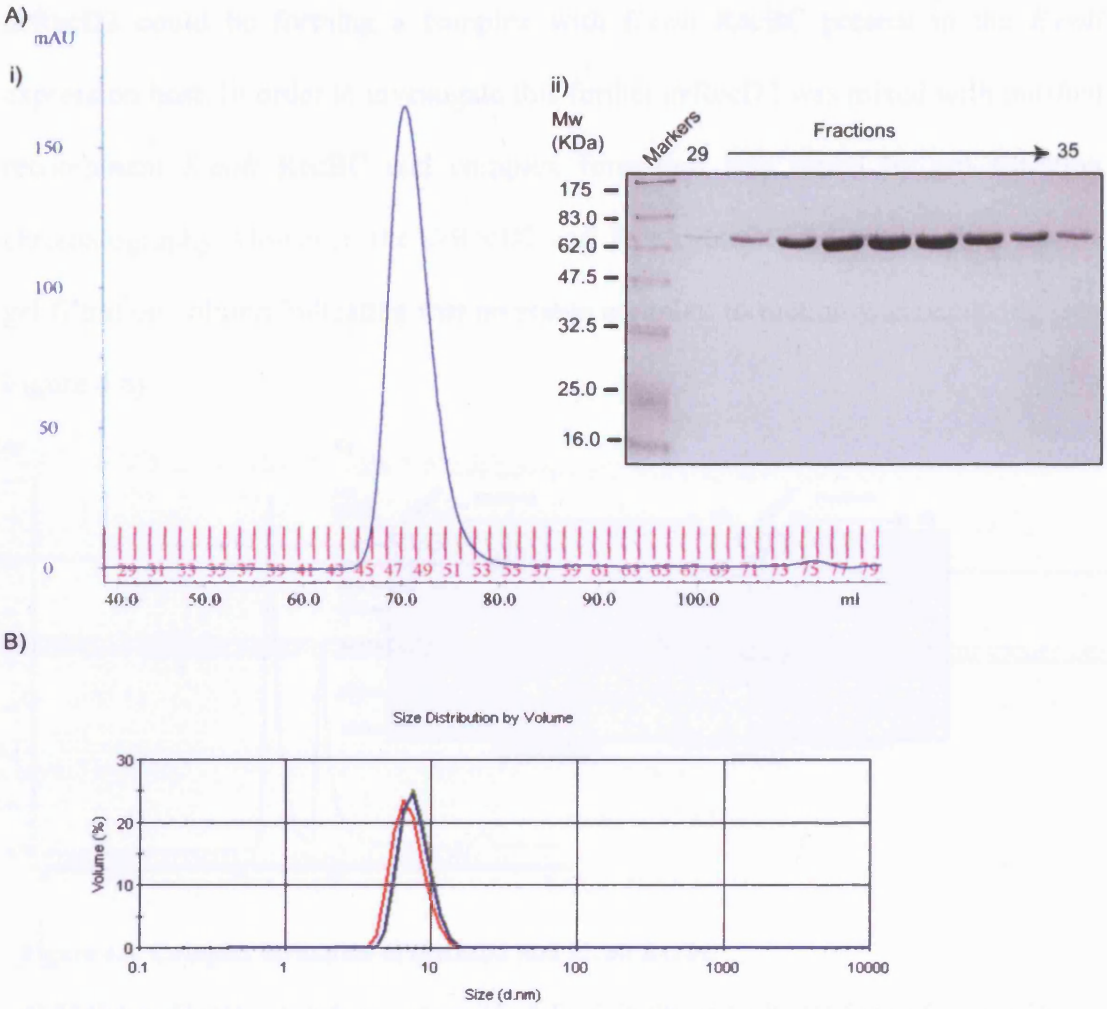

Figure 4.7. Gel filtration and dynamic light scattering analysis of *dr*RecD2

A) i) Elution profile measuring $A_{280\ nm}$ of *dr*RecD2 from a Superdex 200 column. ii) A 12 % SDS-PAGE analysis of the peak fractions eluted.

B) A plot of % volume against size of particle derived from DLS analysis of *dr*RecD2. The % volume is a normalised value derived from the intensity of light scattered by the particles and the mass of the particles.

4.2.4 Possible complex formation between *Deinococcus radiodurans* RecD2 and *E.coli* RecBC

Other contaminants were sometimes observed during the Mono Q step, eluting at a higher sodium chloride concentration than the main peak of *dr*RecD2 (see Figure 4.6 B, fraction 48). These were identified as *E.coli* RecB and RecC by peptide mass mapping, importantly a small amount of *dr*RecD2 was also confirmed by peptide mass mapping to co-elute with the *E.coli* RecBC. This raised the possibility that

*dr*RecD2 could be forming a complex with *E.coli* RecBC present in the *E.coli* expression host. In order to investigate this further *dr*RecD2 was mixed with purified recombinant *E.coli* RecBC and complex formation was tested by gel filtration chromatography. However, the *dr*RecD2 and *E.coli* RecBC did not co-elute from a gel filtration column indicating that no stable complex formation was occurring (see Figure 4.8).

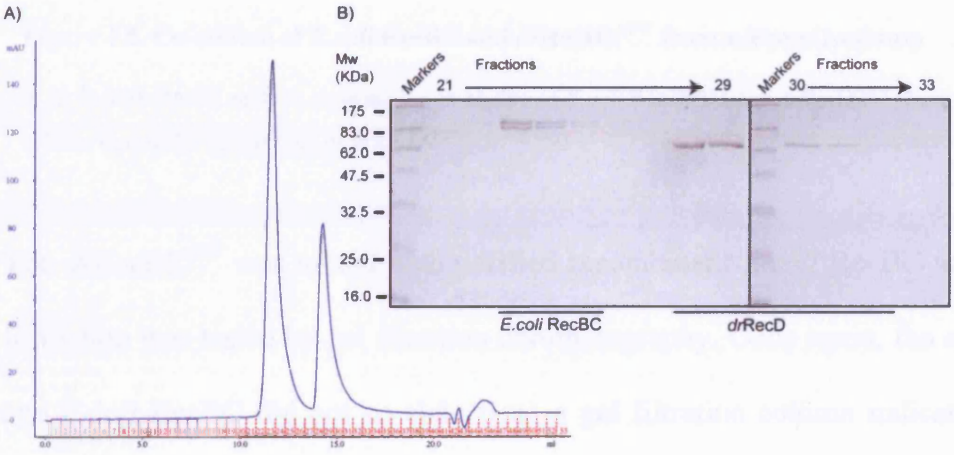

Figure 4.8. Complex formation of *dr*RecD2 and *E.coli* RecBC

A) Elution profile measuring $A_{280\ nm}$ of pre-mixed *E.coli* RecBC and *dr*RecD2 from a Superdex 75 column. B) 12 % SDS-PAGE analysis of the peak fractions eluted from the column.
E.coli RecBC and *dr*RecD2 do not co-elute.

The RecD2 enzymes have an extended N-terminus compared to the RecD1 enzymes. One of the *dr*RecD2 mutants described in this chapter is missing this N-terminal region (*dr*RecD2$^{\Delta NT}$) and is therefore a comparable size to *E.coli* RecD (see section 4.4.3.2). The purification of this mutant is the same as for the wild-type *dr*RecD2, and *E.coli* RecBCD1 were seen to be contaminants that were removed on the MonoQ column with a small amount of co-eluting *dr*RecD2$^{\Delta NT}$ (see Figure 4.9) away from the main peak of purified *dr*RecD2$^{\Delta NT}$.

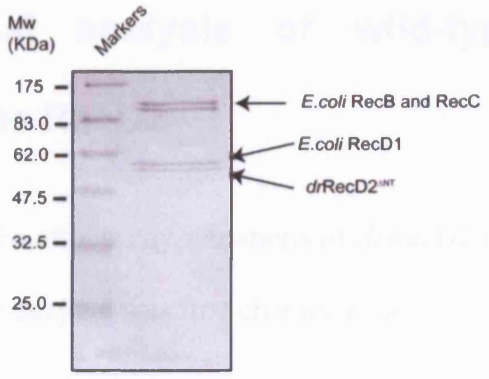

Figure 4.9. Co-elution of *E.coli* RecBC and *dr*RecD2$^{\Delta NT}$ from a MonoQ column

A 12 % SDS-PAGE analysis showing the co-elution of *E.coli* RecBCD1 and *dr*RecD2$^{\Delta NT}$ from a MonoQ column during the purification of *dr*RecD2$^{\Delta NT}$.

The *dr*RecD2$^{\Delta NT}$ was mixed with purified recombinant *E.coli* RecBC and complex formation was tested by gel filtration chromatography. Once again, the *dr*RecD2$^{\Delta NT}$ and *E.coli* RecBC did not co-elute from a gel filtration column indicating that no stable complex formation was occurring (see Figure 4.10).

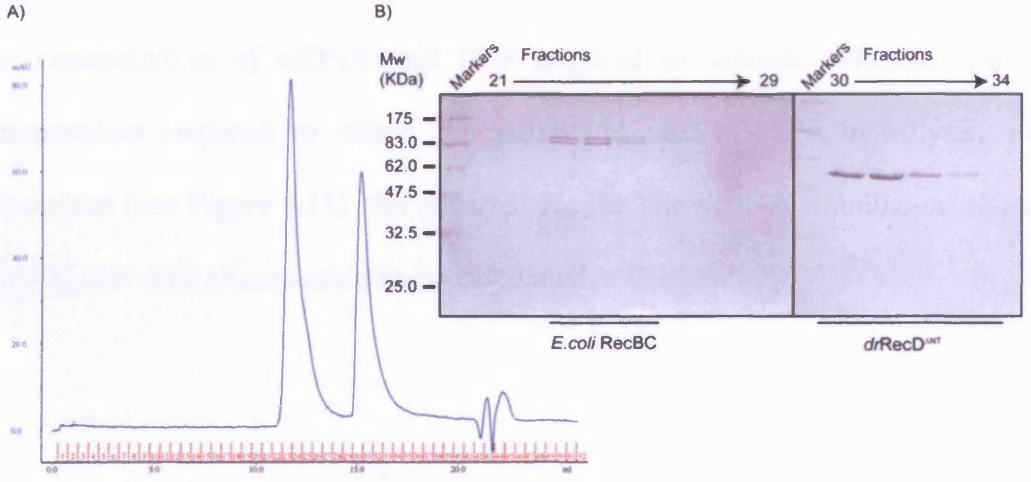

Figure 4.10. Complex formation of *dr*RecD2$^{\Delta NT}$ and *E.coli* RecBC

A) Elution profile measuring $A_{280\ nm}$ of pre-mixed *E.coli* RecBC and *dr*RecD2$^{\Delta NT}$ from a Superdex 75 column. B)A 12 % SDS-PAGE analysis of the peak fractions eluted from the column. *E.coli* RecBC and *dr*RecD2$^{\Delta NT}$ do not co-elute.

4.3 Biochemical analysis of wild-type *Deinococcus radiodurans* RecD2

In order to establish the effects any mutations of *dr*RecD2 might have on its helicase activity, the wild-type enzyme was first characterised.

4.3.1 ATP hydrolysis activity

A pdT60 oligonucleotide was used for this assay in order to make sure that the DNA was of sufficient length to cover the DNA binding sites required to stimulate activity. Isothermal calorimetry (ITC) data show that *dr*RecD2 binds well to pdT substrates longer than 15 nucleotides (these data are described later in section 4.3.2.1), suggesting that all the DNA binding sites of *dr*RecD2 are occupied.

The concentration of ssDNA and ATP required to saturate *dr*RecD2, i.e. the concentration required to obtain the maximum rate of ATP hydrolysis, were determined (see Figure 4.11). An apparent K_m for 'the ssDNA stimulation' (K_{DNA}), and a K_m for ATP (K_{ATP}) can then be calculated.

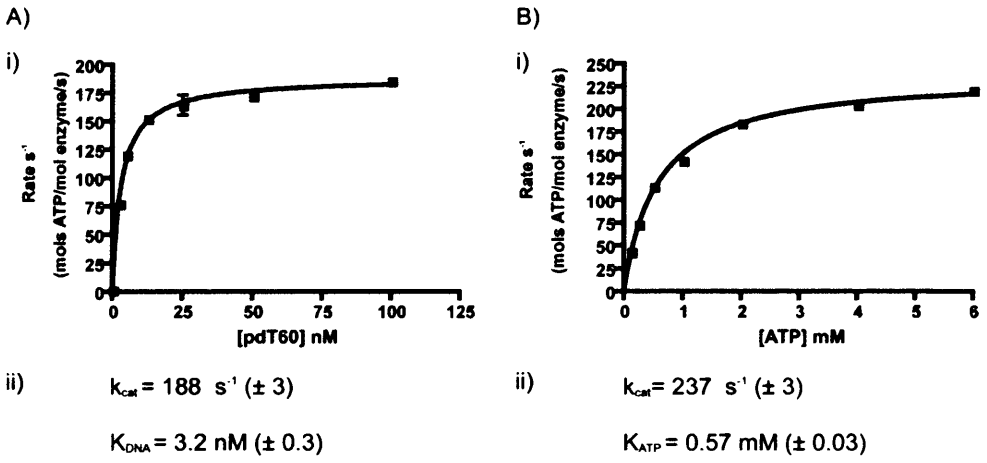

Figure 4.11. ATP hydrolysis activity of *dr*RecD2

A) i) A plot of moles of ATP hydrolysed per mole of *dr*RecD2 per second against pdT60 concentration in order to calculate ii) a k_{cat} and K_{DNA} for the reaction in the presence of saturating (4 mM) ATP.

B) i) A plot of moles of ATP hydrolysed per mole of *dr*RecD2 per second against ATP concentration in order to calculate ii) k_{cat} and K_{ATP} for the reaction in the presence of saturating (100 nM) pdT60.

A more accurate apparent K_m for DNA (K_{DNA2}) was calculated by determining and then plotting the apparent k_{cat} (k_{cat_app}) for various concentrations of pdT60 (see Figure 4.12).

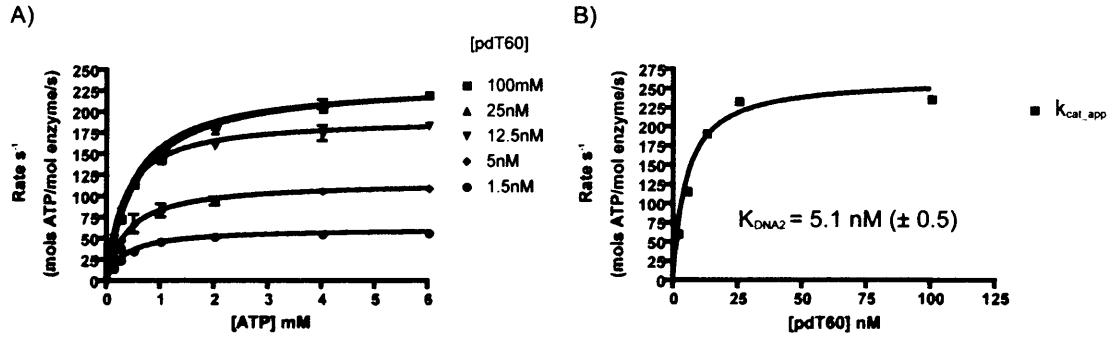

Figure 4.12. Determination of K_{DNA2}

A) A plot of moles of ATP hydrolysed per mole of *dr*RecD2 per second against ATP concentration in order to calculate k_{cat_app} for the reactions at different pdT60 concentrations.

B) A plot of k_{cat_app}, as determined in A), against pdT60 concentration in order to calculate K_{DNA2}.

4.3.2 DNA binding activity

4.3.2.1 Isothermal calorimetry (ITC)

The binding of *dr*RecD2 to ssDNA was investigated in order to characterise the K_d of *dr*RecD2 for ssDNA and the length of ssDNA that could be bound, indicating the length of ssDNA required to fill all the ssDNA-binding pockets of *dr*RecD2.

ITC provides:

i) The stoichiometry of binding (*N*), i.e. the number of molecules of *dr*RecD2 bound to one DNA substrate molecule.

ii) A directly measured association constant of the protein for the DNA substrate (K_a), (where K_a = [*dr*RecD2/DNA complex]/[*dr*RecD2][DNA]), that can be used to calculate its inverse, the dissociation constant, K_d.

ITC experiments were carried out using different length pdT ssDNA substrates. The minimum length of pdT that *dr*RecD2 can bind to was determined (see Figure 4.13).

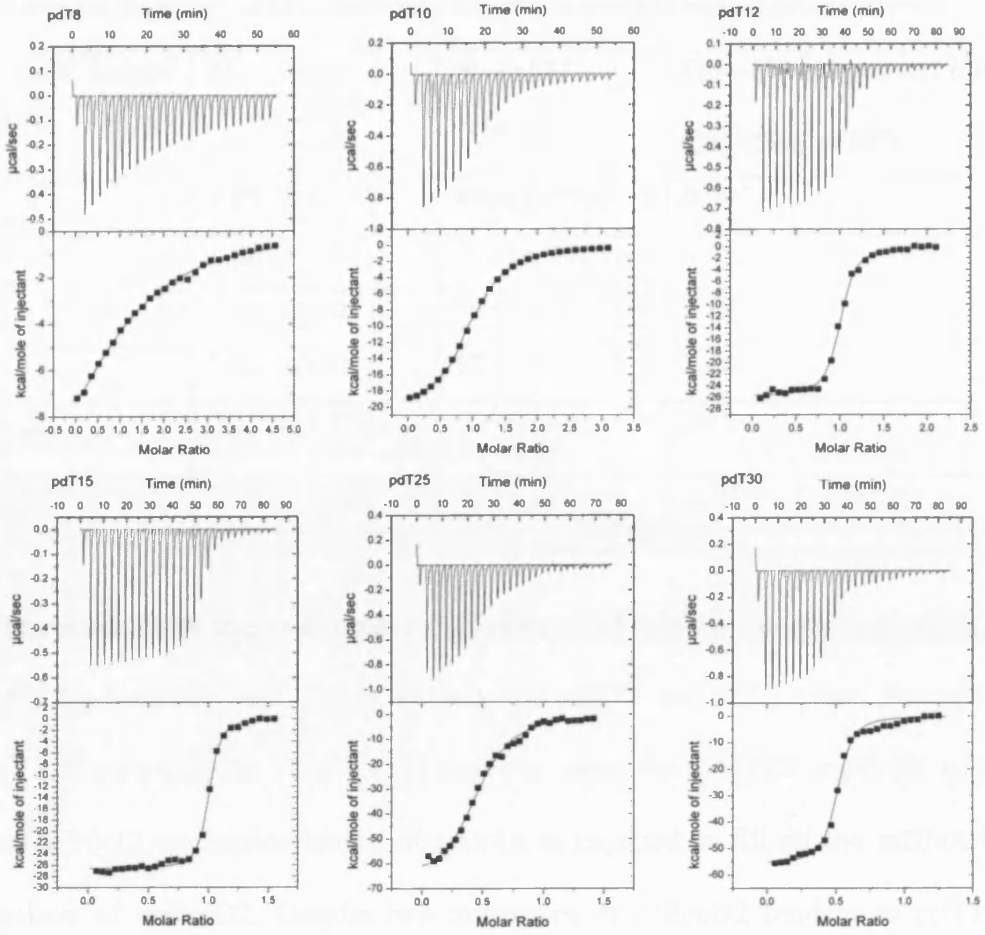

Figure 4.13. ITC analysis of *dr*RecD2 with various pdT substrates

Isothermal calorimetry experiments. Varying lengths of pdT DNA substrates were injected into *dr*RecD2 and the heat of binding measured.

Values for N and K_d were obtained by fitting curves calculated from the one-binding-event model and are summarised in Table 4.2.

Table 4.2. Summary of ITC analysis of *dr*RecD2 binding to various pdT substrates

pdT length	*N*	K_d (nM)	One-binding-event model Chi^2/DoF ($x10^5$)
8	0.89 (± 0.05)	9900 (± 740)	0.09
10	1.00 (± 0.01)	520 (± 77)	0.21
12	1.00 (± 0.01)	56 (± 0.5)	2.84
15	1.00 (± 0.01)	17 (± 3)	1.94
25	2.11 (± 0.03)	400 (± 40)	20.3
30	2.20 (± 0.01)	48 (± 3)	2.70

The results show that one molecule of *dr*RecD2 binds per molecule of pdT8, pdT10, pdT12 or pdT15, with higher affinity for pdT15 and pdT12 than for pdT10 and pdT8. Two molecules of *dr*RecD2 bind per molecule of pdT25 or pdT30, suggesting that a 10-12 nucleotide length of ssDNA is required to fill all the ssDNA-binding pockets of *dr*RecD2. Despite two molecules of *dr*RecD2 binding to pdT30, each *dr*RecD2 molecule binds with the same affinity and so each binding event is indistinguishable. For example, if there is no co-operative binding of the two *dr*RecD2 molecules to the pdT30 molecule, two molecules of *dr*RecD2 binding to one molecule of pdT30 will provide the same experimental ITC data as two molecules of *dr*RecD2 binding to two molecules of pdT15 (see Figure 4.14). The only difference will be the stoichiometry of the number of *dr*RecD2 molecules of bound to one molecule of DNA substrate.

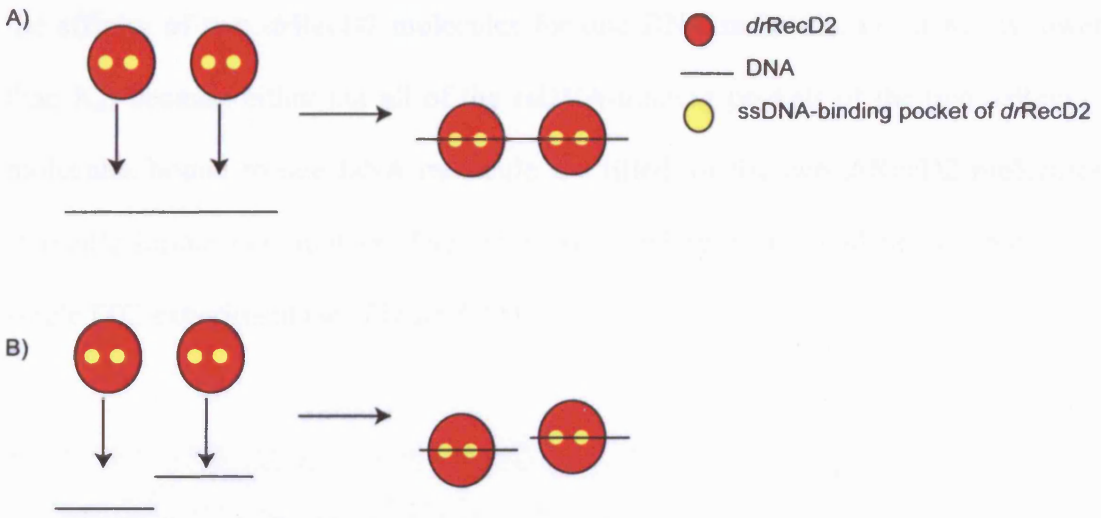

Figure 4.14. Binding events that are indistinguishable by ITC

A) The binding of two *dr*RecD2 molecules to one pdT30 molecule, with no cooperative effects. All the DNA-binding pockets of both *dr*RecD2 molecules are filled.

B) The binding of two *dr*RecD2 molecules to two molecules of pdT15. All the DNA-binding pockets of both *dr*RecD2 molecules are filled.

The heat measured by ITC will be similar for A) and B), only the stoichiometry of the number of molecules *dr*RecD2 bound to one molecule of DNA is different.

To investigate the length of DNA that can accommodate two molecules of *dr*RecD2, ITC experiments were performed using pdT20, pdT21, pdT22, pdT23 and pdT24 ssDNA substrates. In theory this should lead to two binding events being observed (see Figure 4.15). The ITC experiment is carried out by injecting small amounts of DNA substrate into a cell containing *dr*RecD2 and measuring the heat of binding. In these experiments 30 injections were used. After the first injections of DNA, *dr*RecD2 is in excess of DNA. After the last injections of the DNA substrate, the DNA is in excess of the *dr*RecD2. When the *dr*RecD2 is in excess of the DNA, two molecules of *dr*RecD2 may bind to one DNA molecule with the dissociation constant, K_{d1}. When the DNA is in excess of the *dr*RecD2 one of the two *dr*RecD2 molecules bound to one DNA molecule may dissociate, and then bind to another DNA molecule with the dissociation constant, K_{d2}. This rearrangement will only occur if the affinity of one *dr*RecD2 molecule for one DNA molecule is higher than

the affinity of two *dr*RecD2 molecules for one DNA molecule, i.e. if K_{d2} is lower than K_{d1} because either not all of the ssDNA-binding pockets of the two *dr*RecD2 molecules bound to one DNA molecule are filled, or the two *dr*RecD2 molecules sterically hinder one another. Therefore, two binding events will be observed in a single ITC experiment (see Figure 4.15).

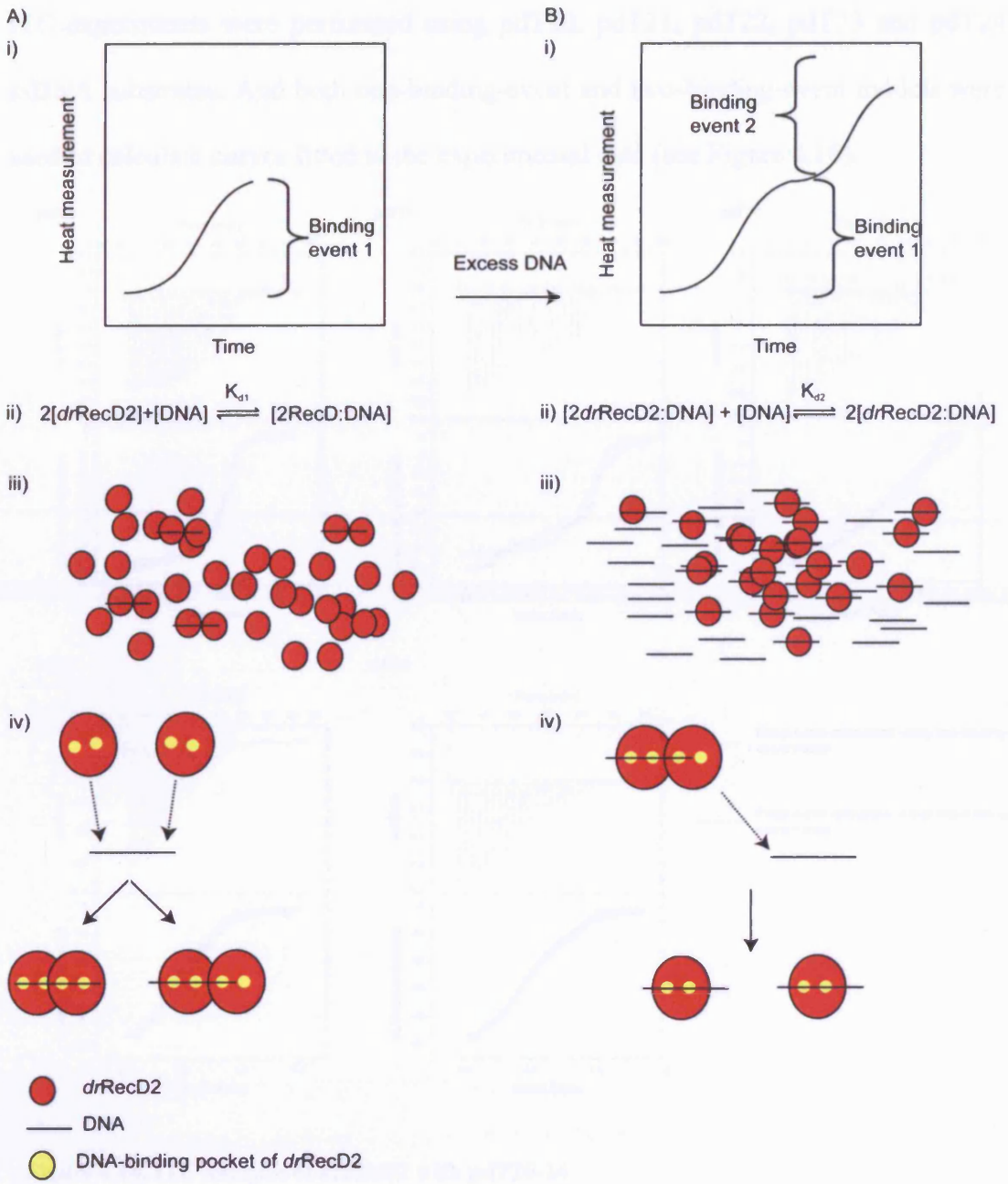

Figure 4.15. Two equilibrium constants in one ITC experiment

A) After the first injections *dr*RecD2 is in excess of the DNA substrate. i) A representation of an ITC binding curve of the equilibrium in ii) when iii) *dr*RecD2 (red) is in excess of DNA (black) and iv) the DNA-binding pockets of one molecule of *dr*RecD2 are not all filled and/or there is steric hindrance between the two *dr*RecD2 molecules.

B) After the last injections the DNA substrate is in excess of *dr*RecD2 i) A representation of an ITC binding curve of the equilibrium in ii) when iii) DNA (black) is in excess of *dr*RecD2 (red) and iv) the DNA-binding pockets of *dr*RecD2 are all filled.

ITC experiments were performed using pdT20, pdT21, pdT22, pdT23 and pdT24 ssDNA substrates. And both one-binding-event and two-binding-event models were used to calculate curves fitted to the experimental data (see Figure 4.16).

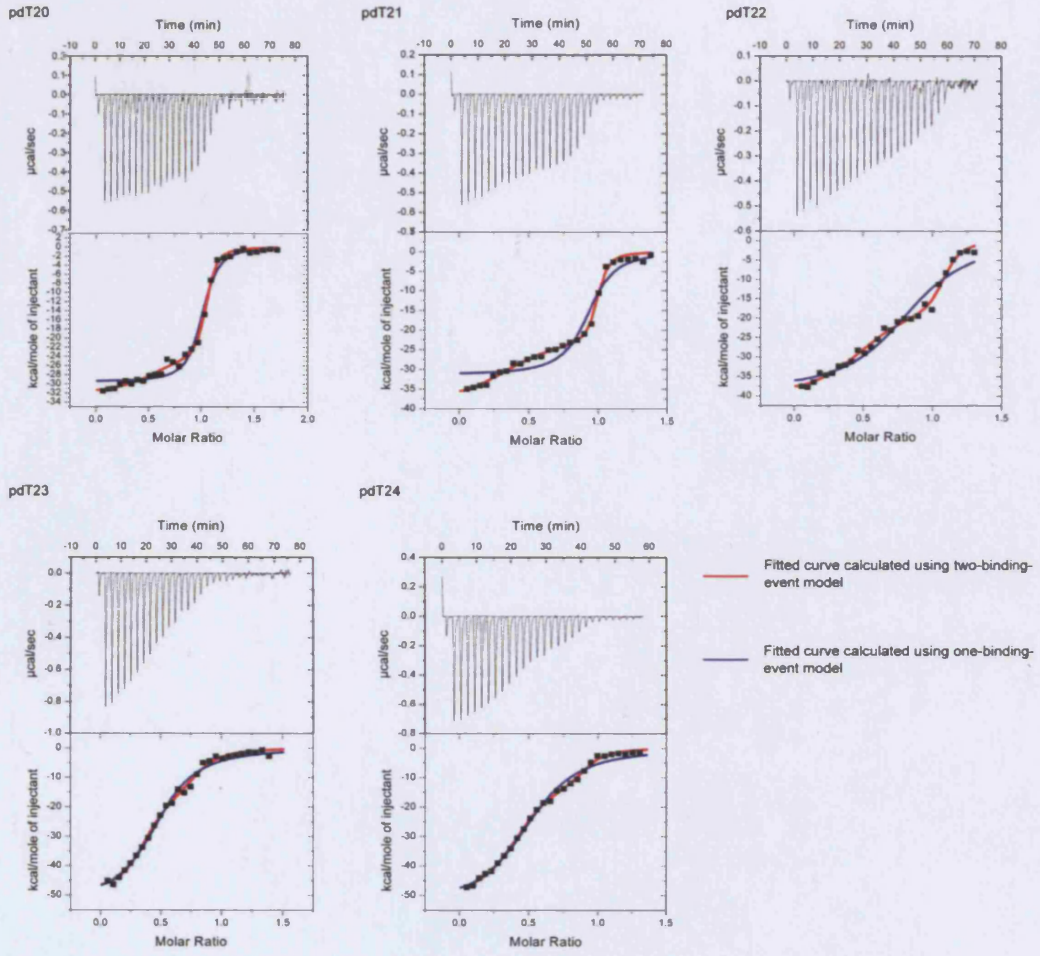

Figure 4.16. ITC analysis of *dr*RecD2 with pdT20-24

Isothermal calorimetry experiments injecting pdT20, pdT21, pdT22, pdT23 or pdT24 into *dr*RecD2.

As well as judging by eye whether the one-binding-event model or the two-binding-event model best fits the data, the fit of the curve to the data can be quantified using the statistical value of Chi^2/DoF. Chi^2 is the sum of the squares of the deviations of the fitted curve from the experimental points and DoF (degree of freedom) is the total number of experimental points used in the fitting minus the total number of 4adjustable parameters. A low Chi^2/DoF value suggests a good fit of the calculated

curve to the experimental data, a high Chi^2/DoF value indicates a bad fit. The Chi^2/DoF values of fitting the curves calculated from the either the one-binding-event or the two-binding-event models to the experimental data of *dr*RecD2 binding to pdT20-24 are listed in Table 4.3.

Table 4.3. Chi^2/DoF of fitting one-binding-event or two-binding-event models to the ITC data of *dr*RecD2 binding to pdT20-24

pdT length	One-binding-event model Chi^2/DoF (x 10^5)	Two-binding-event model Chi^2/DoF (x 10^5)
20	19.3	5.30
21	63.6	6.36
22	54.7	10.9
23	21.0	12.6
24	20.5	4.31

The two-event binding model, judging either by eye or mathematically, best describes *dr*RecD2 binding to pdT20-24. N and K_d were calculated for both binding events with the first and second event denoted as N_1, K_{d1} and N_2, K_{d2} respectively. The results are summarised in Table 4.4.

Table 4.4. Summary of ITC analysis of *dr*RecD2 with pdT20-25

pdT length	N_1	K_{d1} (nM)	N_2	K_{d2} (nM)
20	2.0 (±0.09)	28 (±15)	1.0 (±0.03)	2.4 (±1.4)
21	1.6 (±0.01)	19 (±4)	1.0 (± 0.01)	3.8 (±1.8)
22	1.7 (±0.01)	46 (±9)	1.0 (±0.05)	7.0 (±3.5)
23	2.3 (±0.15)	260 (±100)	1.3 (±0.1)	19 (±8)
24	2.0 (±0.06)	98 (±31)	1.1 (±0.04)	8.6 (±2.5)

As is necessary for the bipartite binding curve to be observed K_{d2} is a lower value than K_{d1}.

The two-binding-event model cannot be used to analyse the ITC data of *dr*RecD2 binding to pdT25 or pdT30 because no result is converged upon using iterative fitting of the calculated curve.

4.3.2.2 DNA binding assays

For comparison of wild-type *dr*RecD2 with mutant *dr*RecD2 proteins a DNA binding assay was used. The binding of *dr*RecD2 to a 5'-tailed duplex, ssDNA and dsDNA was investigated using this technique. The 5'-tailed duplex comprised 20bp duplex DNA with a 5' pdT12 tail. This was chosen for use in both the DNA binding and helicase assay so that the results could be compared. The ssDNA was pdT15, this was chosen because the ITC data shows that one *dr*RecD2 molecule binds to this length of DNA making the DNA binding assays easier to interpret. The dsDNA substrate comprised the duplex region of the 5'-tailed duplex substrate so that the results could be compared. The K_d of *dr*RecD2 for the DNA substrate was determined using DNA binding assays (see Figure 4.17).

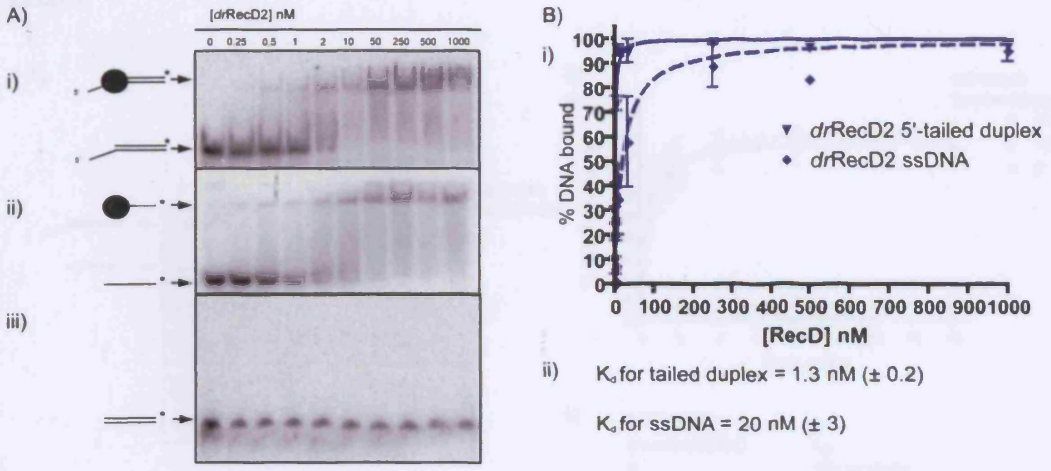

Figure 4.17. DNA binding activity of *dr*RecD2

A) Autoradiograph of a 9 % native PAGE analysis of the DNA binding assays of *dr*RecD2 with i) the 5'-tailed duplex, ii) the ssDNA and iii) the dsDNA substrates. *indicates radio-labelled DNA substrate.
B) Plot of % DNA bound against *dr*RecD2 concentration in order to calculate ii) a K_d of *dr*RecD2 for the tailed duplex and ssDNA substrates.

The results show that *dr*RecD2 cannot bind dsDNA and binds to the 5'-tailed duplex substrate with a higher affinity than to the ssDNA substrate.

4.3.3 Helicase activity

Previous work has shown that *dr*RecD2 and *E.coli* RecD1 have 5'-3' helicase directionality, i.e. the helicases unwind duplex DNA that has a 5'-ssDNA tail (Dillingham *et al.* 2003; Wang and Julin 2004). A helicase assay was performed in order to compare wild-type *dr*RecD2 to the mutant *dr*RecD2s. (see Figure 4.18). The substrate used was the same 5'-tailed duplex used in the DNA binding assays. The time taken to unwind 50% of the DNA substrate was determined ($t_{1/2}$), as was the length of the 5'-tail required for activity. It is important to note that $t_{1/2}$ is not a constant and will vary with the concentrations of both the enzyme and the DNA substrate used in the assay. However, all the helicase assays in this thesis use the

same concentration of enzyme and substrate so $t_{1/2}$ can be directly compared between assays.

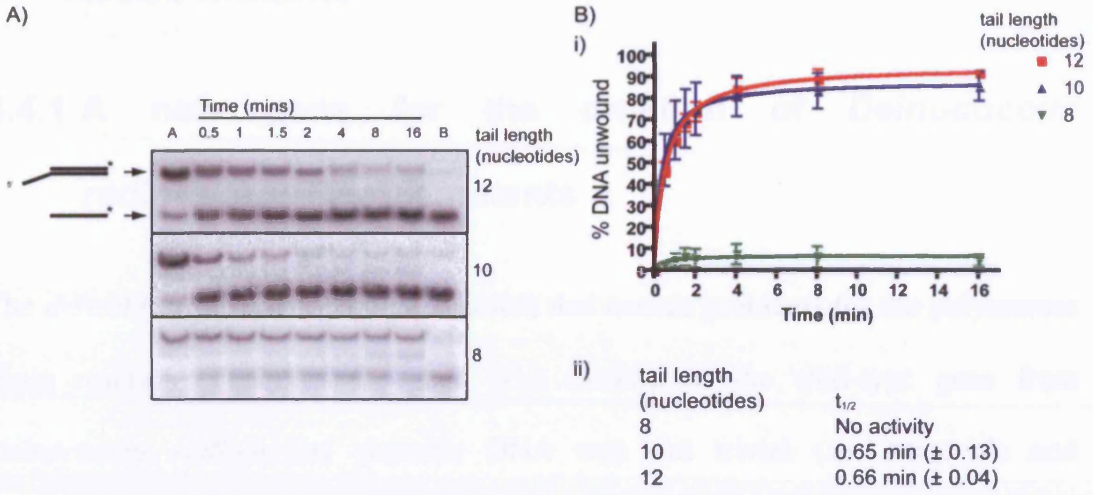

Figure 4.18. Helicase activity of *dr*RecD2

A) Autoradiograph of a 10% SDS-PAGE analysis of the helicase assays of *dr*RecD2. The substrates were all 5'-tailed duplex DNA, the length of the 5' tail varied from 8 to 12 nucleotides. 'A' indicates annealed substrate at time 0 min. 'B' indicates boiled substrate as marker for ssDNA. *indicates radio-labelled DNA. B) Plot of % DNA unwound against time in order to calculate ii) a $t_{1/2}$ for the substrates.

4.4 Biochemical analysis of *Deinococcus radiodurans* RecD2 mutants

4.4.1 A new gene for the creation of *Deinococcus radiodurans* RecD2 mutants

The *dr*RecD2 gene has a high GC content that causes problems for the polymerase chain reaction (PCR) of the gene. The cloning of the wild-type gene from *Deinococcus radiodurans* genomic DNA was not trivial (see Materials and Methods). Despite numerous attempts and conditions the mutants could not be cloned using the *dr*RecD2 gene as a template because the replication of DNA step using PCR was unsuccessful. This problem was overcome by redesigning the wild-type gene sequence utilising codons recognised by the most abundant tRNAs present in *E.coli* whilst not changing the amino acid coding sequence. Changing the codons in this way had two consequences. Firstly, the gene sequence was less GC rich than the wild-type gene sequence allowing the replication of DNA using PCR. Secondly, the gene was better translated from mRNA to protein in *E.coli* leading to more protein being produced. The new sequence was designed and synthesised commercially. The new *dr*RecD2 gene (*dr*RecD2$_{new}$) was cloned into an expression vector and underwent expression trials. The gene expressed soluble protein in various cell lines (data not shown). The amount of protein expressed from the *dr*RecD2$_{new}$ gene was more than that from the *dr*RecD2 gene. However, not all of the protein expressed from the *dr*RecD2$_{new}$ was soluble (see Figure 4.19). BL21 *E.coli* was chosen as the expression strain as it grew the fastest, with cultures reaching an O.D$_{600nm}$ of 0.5 – 0.7 AU within three hours compared to six or seven hours for other *E.coli* expression strains.

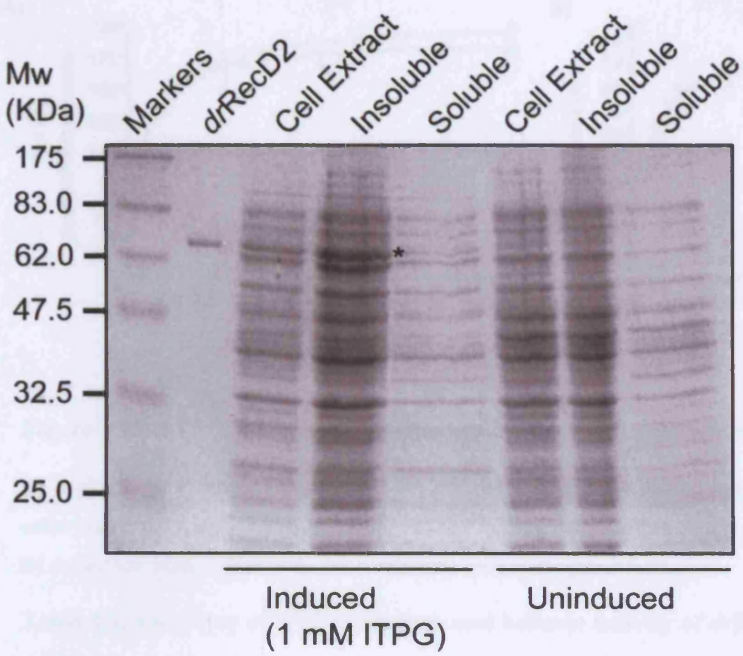

Figure 4.19. Expression of *dr*RecD2 protein from the *dr*RecD2$_{new}$ gene

A 12% SDS-PAGE analysis of the induced expression of soluble *dr*RecD2 protein from the *dr*RecD2$_{new}$ gene. The *E.coli* BL21 with the *dr*RecD2 expression vector was grown to 0.6 O.D$_{600nm}$ and was induced with 1mM IPTG for 3 hours. A control culture was not induced. Purified *dr*RecD2 as previously described was used as a marker for the molecular weight of *dr*RecD2. *indicates soluble *dr*RecD2 protein expressed from the *dr*RecD2$_{new}$ gene.

The *dr*RecD2 protein expressed from *dr*RecD2$_{new}$ gene was purified in the same way as *dr*RecD2 protein expressed from the wild-type *dr*RecD2 gene (see section 4.2.3).

The protein expressed from the *dr*RecD2$_{new}$ gene had the same enzymatic activity as the protein expressed from the wild-type *dr*RecD2 gene (see Figure 4.20 and Table 4.5).

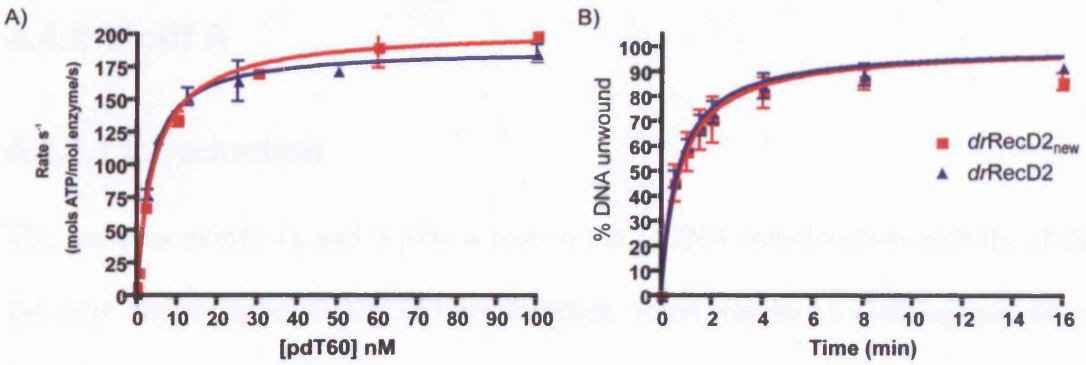

Figure 4.20. ATP hydrolysis and helicase activity of *dr*RecD2 protein from *dr*RecD2$_{new}$ gene

A) A plot of moles of ATP hydrolysed per mole of *dr*RecD2 per second against pdT60 concentration in order to calculate a k_{cat} and K_{DNA} for the reaction in the presence of saturating (4 mM) ATP.

B) A helicase assay, a plot of % DNA substrate unwound against time.

Table 4.5. Summary of ATP hydrolysis and helicase activity of *dr*RecD2

	ATP hydrolysis K_{DNA} (nM)	**ATP hydrolysis k_{cat} (s^{-1})**	**Helicase $t_{1/2}$ (min)**
*dr*RecD2	3.2 (± 0.3)	188 (± 3)	0.66 (± 0.04)
*dr*RecD2$_{new}$	4.5 (± 0.5)	203 (± 4)	0.75 (± 0.06)

4.4.2 Motif A

4.4.2.1 Introduction

The helicase motifs 1a and 3 play a role in the ssDNA translocation activity of the *Bacillus stearothermophilus* SF1Aα helicase, PcrA (*bs*PcrA) (Dillingham *et al.* 2001; Velankar *et al.* 1999). Phenylalanine 64 of motif 1a moves during the catalytic cycle of ATP hydrolysis by *bs*PcrA, opening and closing a hydrophobic DNA-binding pocket. The opening and closing of the hydrophobic DNA-binding pocket forces the bases of ssDNA bound to *bs*PcrA to move (see Figure 4.21).

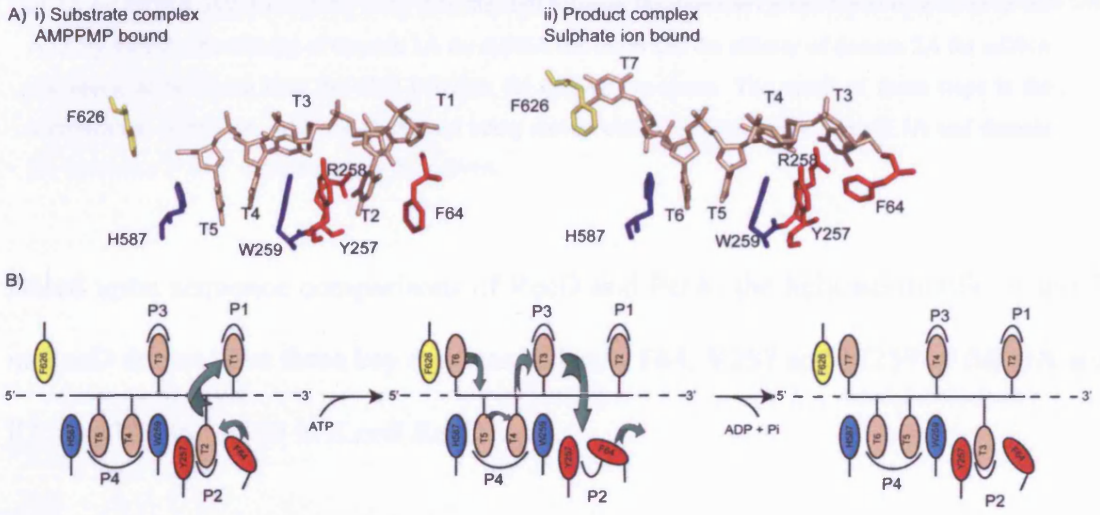

Figure 4.21. Helicase motifs 1a and 3 play a role in the ssDNA translocation by *bs*PcrA.

The DNA is coloured brown and the residues of binding pockets P2 and P4 are coloured red and blue respectively. F626 is coloured yellow. F64 of helicase motif 1a is shown as are Y257 and W259 of helicase motif 3.

A) Structural detail of the residues that comprise the ssDNA-binding sites of *bs*PcrA. i) The substrate complex with AMPPNP is bound. ii) The product complex with a sulphate ion bound that mimics a phosphate.

B) A schematic model of ssDNA-binding pockets of *bs*PcrA at different stages of the catalytic cycle of ATP hydrolysis. Upon ATP binding to PcrA F64 moves into pocket P2 displacing the base T2 into pocket P1, and T1 is pushed out of the protein and slides over domain 1A. This is accompanied by a closing of the cleft between domains 1A and 2A. Upon ATP hydrolysis F64 moves out of pocket 2 andT3 flips into the pocket. The movement of T3 forces T4 to flip into pocket P3, which in turn forces T5 to move and T6 to flip into pocket P4. This is accompanied by a closing of the cleft between domains 1A and 2A.

This figure is based upon figures by Velankar *et al.* (Velankar *et al.* 1999).

The forced movement of the DNA bases and the opening and closing of the cleft between domain 1A and domain 2A results in DNA translocation by an inchworm mechanism (see Figure 4.22)

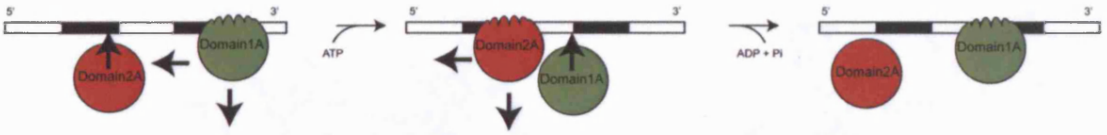

Figure 4.22. Translocation of ssDNA by domain 1A and domain 2A of *bs*PcrA

Upon ATP binding to *bs*PcrA the affinity of domain 1A for ssDNA reduces while the affinity of domain of 2A for ssDNA increases, at the same time the cleft between the domains closes. Upon hydrolysis of the ATP by *bs*PcrA, the affinity of domain 1A for ssDNA increases and the affinity of domain 2A for ssDNA decreases, at the same time the cleft between the domains re-opens. The result of these steps is the translocation of ssDNA, with the 3' tail end being moved out of *bs*PcrA while domain 1A and domain 2A move in a 3' to 5' direction along the ssDNA.

Based upon sequence comparisons of RecD and PcrA, the helicase motifs 1a and 3 in RecD do not have three key residues of PcrA. F64, Y257 and W259 of *bs*PcrA are P205, S303 and E305 in *E.coli* RecD1.

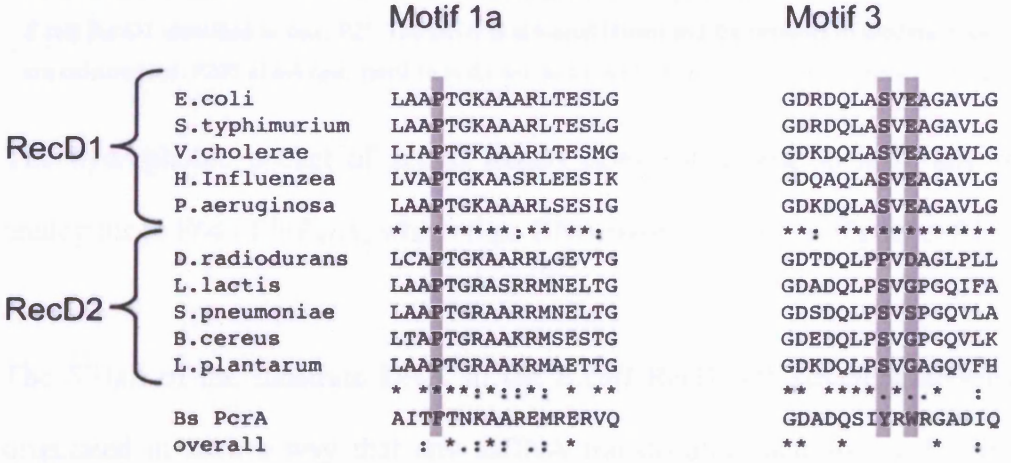

Figure 4.23. A sequence comparison of the helicase motifs 1a and 3 in RecD and *bs*PcrA

F64, Y257 and W259 of are indicated in grey, as are the residues in an equivalent position in the motifs of the RecD family of enzymes.

An updated structure of *E.coli* RecBCD1 (personal communication, Wigley, D.) shows that a hydrophobic pocket is present in *E.coli* RecD1, of which P205 is a part (see Figure 4.24).

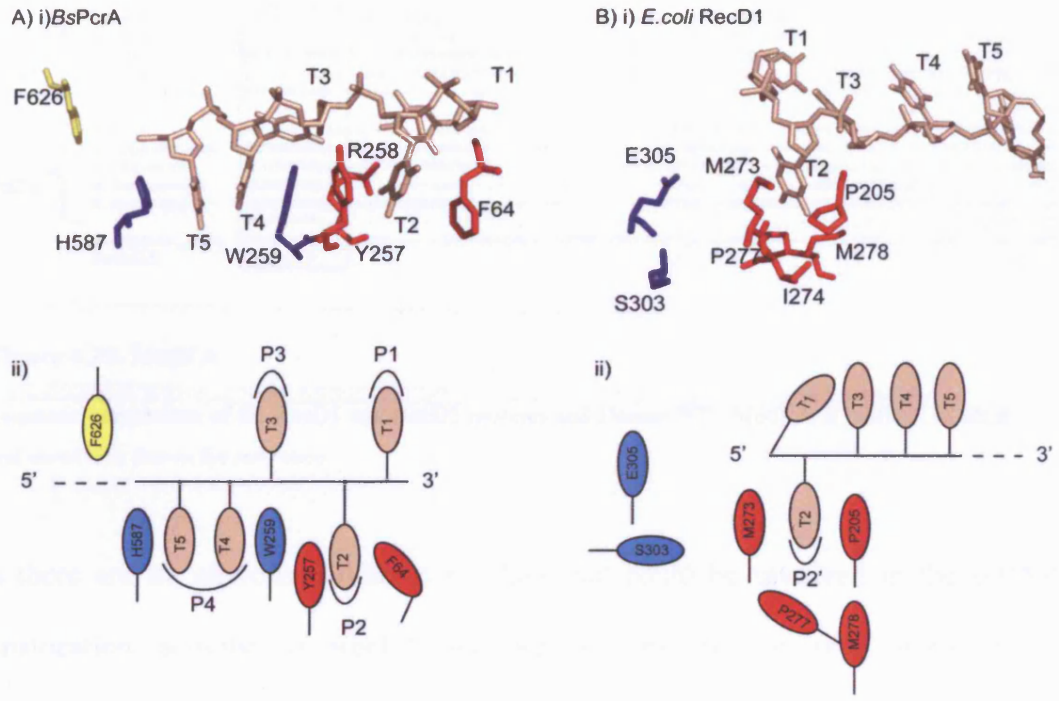

Figure 4.24. T he ssDNA-binding pockets of *bs*PcrA and *E.coli* RecD1

A) i) Structural detail and ii) a schematic of the residues that comprise the ssDNA-binding sites of *bs*PcrA. The DNA is coloured brown and the residues of binding pockets P2 and P4 are coloured red and blue respectively. F64 of helicase motif 1a is shown as are Y257 and W259 of helicase motif 3.

B) Structural detail and ii) a schematic of the residues that comprise the only ssDNA-binding site of *E.coli* RecD1 identified to date, P2'. The DNA is coloured brown and the residues of binding pocket P2' are coloured red. P205 of helicase motif 1a is shown, as are S303 and E305 of helicase motif 3 (blue).

The hydrophobic pocket of *E.coli* RecD1 does not appear to have any residues analogous to F64 of *bs*PcrA, which flips DNA bases in and out of pocket P2.

The 5'-tail of the substrate DNA in the *E.coli* RecBCD1 structure appears to be orientated in such a way that any ssDNA translocation activity by *E.coli* RecD1 would push the 5' end of the ssDNA toward domain 2B of *E.coli* RecD1. However, the 2B domain of *E.coli* RecD1 is disordered in the structure of *E.coli* RecBCD1, so it is not possible to say if domain 2B interacts with the ssDNA substrate. Sequence

comparisons of RecD show a conserved motif (motif A) in domain 2B of types 1 and 2 RecD proteins, that has also been reported to exist in the SF1Bα helicase Pif1 (Zhang *et al.* 2006a)(see Figure 4.25).

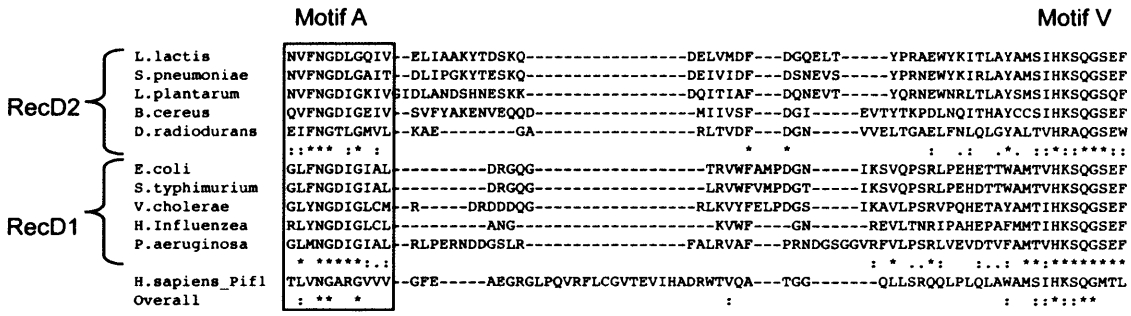

Figure 4.25. Motif A

Sequence comparison of the RecD1 and RecD2 proteins and Human Pif1. Motif A is outlined in black and motif V is shown for reference.

As there are no obvious candidate residues that could be involved in the ssDNA translocation activity of RecD1 and RecD2 enzymes, it was reasoned the phenylalanine in motif A of the 2B domain might interact with ssDNA in a similar fashion to phenylalanine 64 in *bs*PcrA. In order to investigate the role of the conserved motif A, phenylalanine 635 was mutated to an alanine in *dr*RecD2 to make *dr*RecD2^{F635A}.

4.4.2.2 Biochemical analysis of *dr*RecD2^{F635A}

Phenylalanine 635 of *dr*RecD2 was mutated to an alanine to produce *dr*RecD2^{F635A} and the protein was expressed and purified (see Figure 4.26).

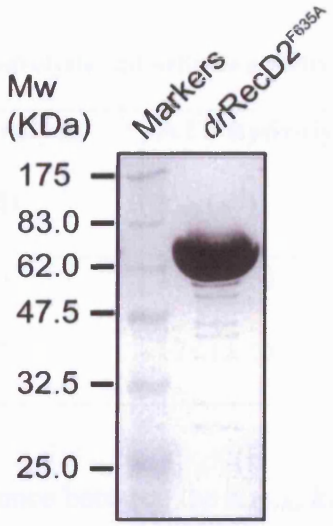

Figure 4.26. Purified *dr*RecD2^{F635A}

A 12% SDS-PAGE analysis of the purity of *dr*RecD2^{F635A}.

ATP hydrolysis and helicase assays of *dr*RecD2^{F635A} were performed (see Figure 4.27 and Table 4.6).

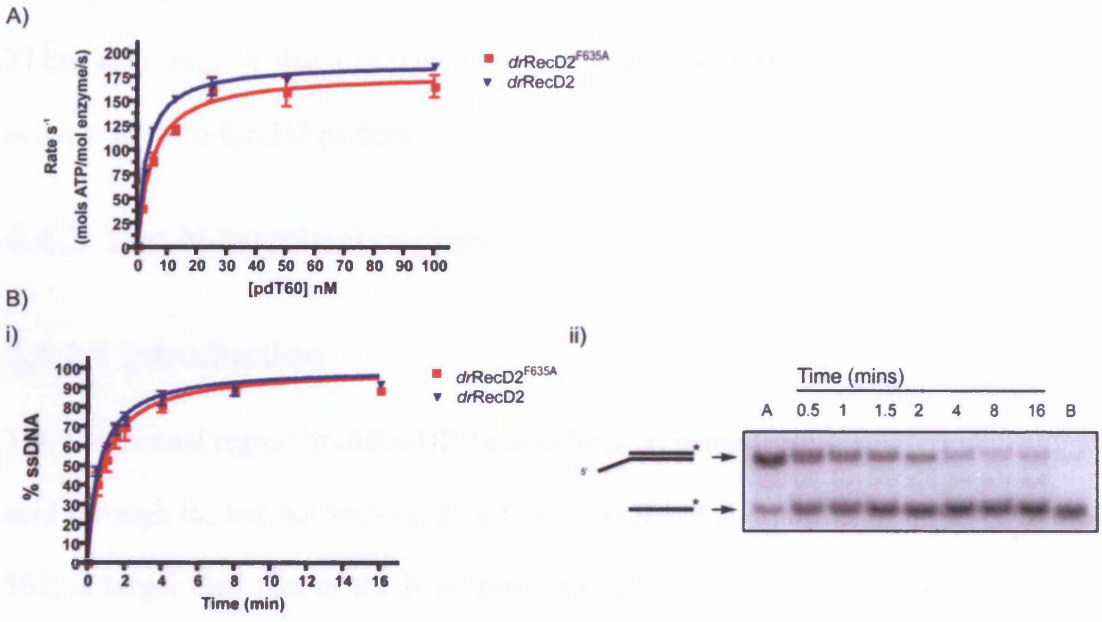

Figure 4.27. ATP hydrolysis and helicase activity of *dr*RecD2^{F635A}

A) ATP hydrolysis activity. A plot of moles of ATP hydrolysed per mole of *dr*RecD2 or *dr*RecD2^{F635A} per second against pdT60 concentration in order to calculate a k_{cat} and K_{DNA} for the reaction in the presence of saturating (4 mM) ATP.

B) Helicase activity. i) A plot of % DNA substrate unwound against time. ii) Autoradiograph of a 10% SDS-PAGE analysis of a helicase assay of *dr*RecD2^{F635A}. 'A' indicates annealed substrate at time 0 min. 'B' indicates boiled substrate as marker for ssDNA. *indicates radio-labelled DNA.

Table 4.6. Summary of ATP hydrolysis and helicase activity of *dr*RecD2^{F635A}

	ATP hydrolysis K_{DNA} (nM)	**ATP hydrolysis k_{cat} (s^{-1})**	**Helicase $t_{1/2}$ (min)**
*dr*RecD2	3.2 (± 0.3)	188 (± 3)	0.66 (± 0.04)
*dr*RecD2^{F635A}	5.0 (± 0.6)	178 (± 4)	0.85 (± 0.08)

There is no significant difference between the K_{DNA}, k_{cat} and $t_{1/2}$ of *dr*RecD2^{F635A} and those of *dr*RecD2.

DNA binding assays were not performed in this instance. The ATP hydrolysis assay is also an indicator of DNA binding and *dr*RecD2^{F635A} has the same ATP hydrolysis activity as *dr*RecD2. The helicase activity was also not affected by the mutation. These data suggest that phenylalanine 635 has no significant role in the helicase activity of the *dr*RecD2 protein.

4.4.3 The N-terminal region

4.4.3.1 Introduction

The N-terminal region of *dr*RecD2, here defined as being from the N-terminal amino acid through to, but not including, helicase motif 1 and comprising amino acids 1-362, is larger than that of the N-terminal domain 1 of *E.coli* RecD1 that comprises amino acids 1-172. The type 2 RecD enzymes characteristically have a larger N-terminal region than the type 1 RecD helicases (see Figure 4.28).

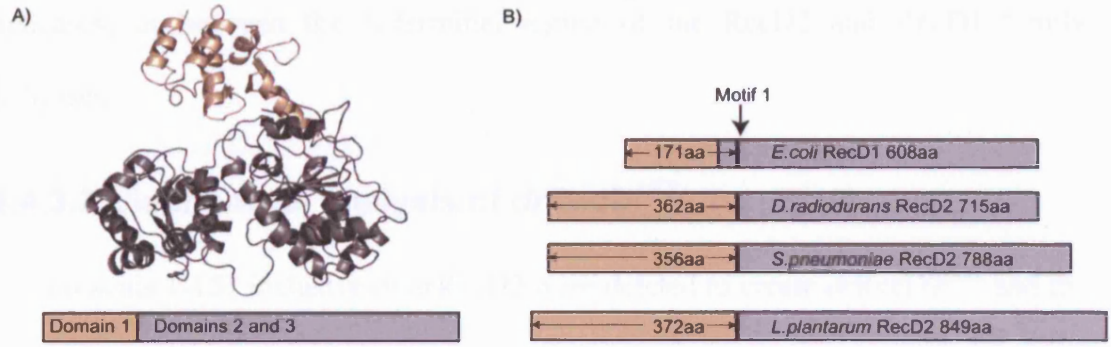

Figure 4.28. A comparison of the N-terminal domains of RecD2 and RecD1 enzymes

A) The structure and a schematic of *E.coli* RecD1. The N-terminal domain 1 is coloured in brown and the rest of the protein in grey.

B) A schematic comparison of the N-terminal domain 1 of *E.coli* RecD1 coloured in brown with the N-terminal of three examples of RecD2 enzymes also coloured in brown.

Domain 1 of *E.coli* RecD1 is responsible for binding to *E.coli* RecC in the *E.coli* RecBCD1 complex. This cannot be the function of the N-terminal region of the RecD2 helicases because, by definition, there is no *bona fide* RecC, or RecC homologue available for RecD2 to bind to. It was reasoned that the N-terminal region of the RecD2 helicases could be responsible for disrupting the duplex DNA ahead of the helicase motor domains, similar to the function of the 2B domain of PcrA or the N-terminal domain of RecG (Singleton *et al.* 2001; Soultanas *et al.* 2000; Velankar *et al.* 1999). During the purification of *dr*RecD2 a distinct proteolysic fragment was isolated (see section 4.2.3). The fragment was N-terminally sequenced, which revealed *dr*RecD2 had been proteolysed between tryptophan 150 and serine 151 indicating that this point could be between two distinct regions of secondary structure. It was reasoned that this would be a good point to make an N-terminal truncation as the protein would be stable and soluble and would probably not undergo further proteolysis. A truncated *dr*RecD2 was produced from amino acid 151 to the C-terminal end amino acid, (*dr*RecD2$^{\Delta NT}$). There is no obvious homology between the N-terminal regions of the RecD2 family

helicases, or between the N-terminal region of the RecD2 and RecD1 family helicases.

4.4.3.2 Biochemical analysis of *dr*RecD$^{\Delta NT}$

Amino acids 1-151 inclusive of *dr*RecD2 were deleted to create *dr*RecD2$^{\Delta NT}$ and the protein was expressed and purified (see Figure 4.29).

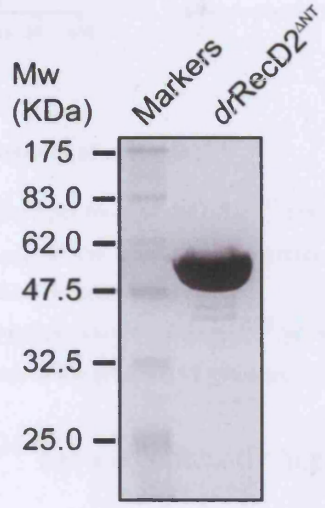

Figure 4.29. Purity of *dr*RecD2$^{\Delta NT}$

A 12% SDS-PAGE analysis of the purity of *dr*RecD2$^{\Delta NT}$.

4.4.3.2.1 ATP hydrolysis activity

The ATP hydrolysis activity of *dr*RecD2$^{\Delta NT}$ was investigated (see Figure 4.30).

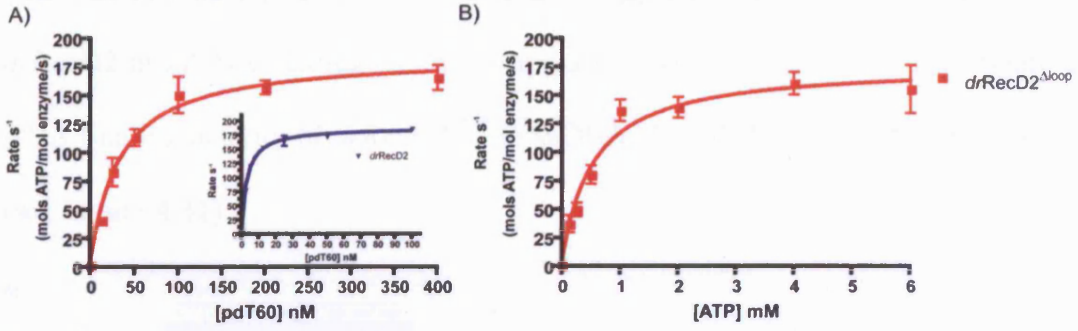

Figure 4.30. ATP hydrolysis activity of *dr*RecD2$^{\Delta NT}$

A) A plot of moles of ATP hydrolysed per mole of *dr*RecD2$^{\Delta NT}$ per second against pdT60 concentration in order to calculate a k_{cat} and K_{DNA} for the reaction in the presence of saturating (4 mM) ATP. The *dr*RecD2 ATP hydrolysis assay is inset for comparison.

B) A plot of moles of ATP hydrolysed per mole of *dr*RecD2$^{\Delta NT}$ per second against ATP concentration in order to calculate K_{ATP} of the reaction in the presence of saturating (3.2 μM) pdT60 DNA substrate.

The results show *dr*RecD2$^{\Delta NT}$ has a significantly higher K_{DNA} than that of wild-type *dr*RecD2, suggesting that *dr*RecD2$^{\Delta NT}$ has a higher K_d for ssDNA than that of wild-type *dr*RecD2. The K_{ATP} of *dr*RecD2$^{\Delta NT}$ was similar to that of wild-type *dr*RecD2, suggesting that ATP binding was unaffected by the mutation (see Table 4.7).

Table 4.7. Summary of ATP hydrolysis activity of *dr*RecD2$^{\Delta NT}$

	K_{DNA} (nM)	k_{cat} (s^{-1})	K_{ATP} (mM)
*dr*RecD2	3.2 (± 0.3)	188 (± 3)	0.57 (±0.03)
*dr*RecD2$^{\Delta NT}$	230 (± 15)	199 (± 3)	0.78 (±0.13)

The k_{cat} of *dr*RecD2$^{\Delta NT}$ was comparable to that of wild-type *dr*RecD2 indicating the ATP hydrolysis activity was not affected by the mutation when enough ssDNA was present to overcome the defect in DNA binding and stimulate the ATP hydrolysis activity of *dr*RecD2$^{\Delta NT}$ (see Table 4.7).

4.4.3.2.2 DNA binding activity

The ATP hydrolysis assay had indicated that *dr*RecD2$^{\Delta NT}$ has a higher K_d for ssDNA than that of wild-type *dr*RecD2. These data suggested a role for the N-terminus of *dr*RecD2 in ssDNA binding. DNA binding assays were carried out to investigate the DNA binding activity of *dr*RecD2$^{\Delta NT}$ to ssDNA, 5'-tailed duplex DNA and dsDNA (see Figure 4.31).

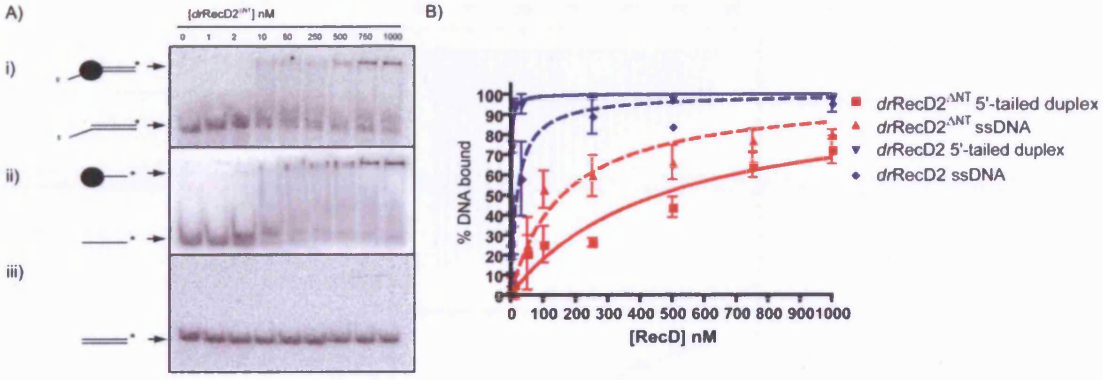

Figure 4.31. DNA binding activity of *dr*RecD2$^{\Delta NT}$

A) Autoradiograph of a 9% native PAGE analysis of the DNA binding assay of *dr*RecD2$^{\Delta NT}$ with i) the 5'-tailed duplex , ii) the ssDNA and, iii) dsDNA substrates. *indicates radio-labelled DNA substrate.
B) A plot of % DNA bound against *dr*RecD2$^{\Delta NT}$ concentration in order to calculate a K_d of *dr*RecD2$^{\Delta NT}$ for the 5'-tailed duplex and ssDNA substrates.

The K_d of *dr*RecD2$^{\Delta NT}$ for the 5'-tailed duplex substrate is approximately 500 times greater than that of wild-type *dr*RecD2 (see Table 4.8).

Table 4.8. Summary of DNA binding activity of *dr*RecD2$^{\Delta NT}$

	K_d for the 5'-tailed duplex substrate (nM)	**K_d for the ssDNA substrate (nM)**	**K_d for the dsDNA substrate (nM)**
*dr*RecD2	1.3 (± 0.2)	20 (± 3)	No binding
*dr*RecD2$^{\Delta NT}$	461 (± 45)	160 (± 16)	No binding

The K_d of *dr*RecD2$^{\Delta NT}$ for the ssDNA substrate is approximately 5 times greater than that of *dr*RecD2, and no binding of *dr*RecD2$^{\Delta NT}$ to dsDNA is observed (see Table

4.8). These data would suggest that the N-terminus of *dr*RecD2 plays a significant role in the binding duplex DNA in the context of a 5'-tailed duplex DNA structure.

4.4.3.2.3 Isothermal calorimetry (ITC)

In order to confirm the defect in ssDNA binding, an ITC experiment was performed with *dr*RecD2$^{\Delta NT}$ and pdT15 (see Figure 4.32).

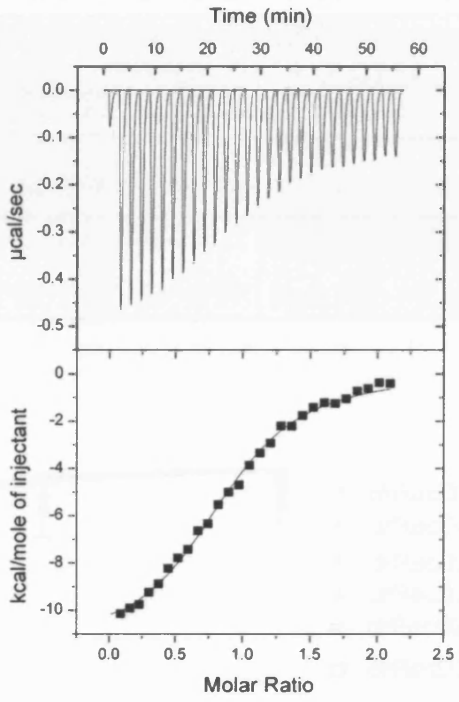

Figure 4.32. Isothermal calorimetry of *dr*RecD2$^{\Delta NT}$ and pdT15

Isothermal calorimetry experiment. The ssDNA substrate, pdT15, was injected into *dr*RecD2$^{\Delta NT}$.

The results as compared with wild-type *dr*RecD2 are summarised in Table 4.9.

Table 4.9. Summary of ITC results of *dr*RecD2$^{\Delta NT}$

	N	**K_d (nM)**
*dr*RecD2	1.0 (±0.01)	17 (±3)
*dr*RecD2$^{\Delta NT}$	1.1 (±0.02)	1170 (±190)

4.4.3.2.4 *Helicase activity*

The helicase activity of *dr*RecD2$^{\Delta NT}$ was investigated using helicase assays (see Figure 4.33). All the helicase assays contained 1 nM *dr*RecD2 or *dr*RecD2$^{\Delta NT}$ and 1 nM DNA substrate.

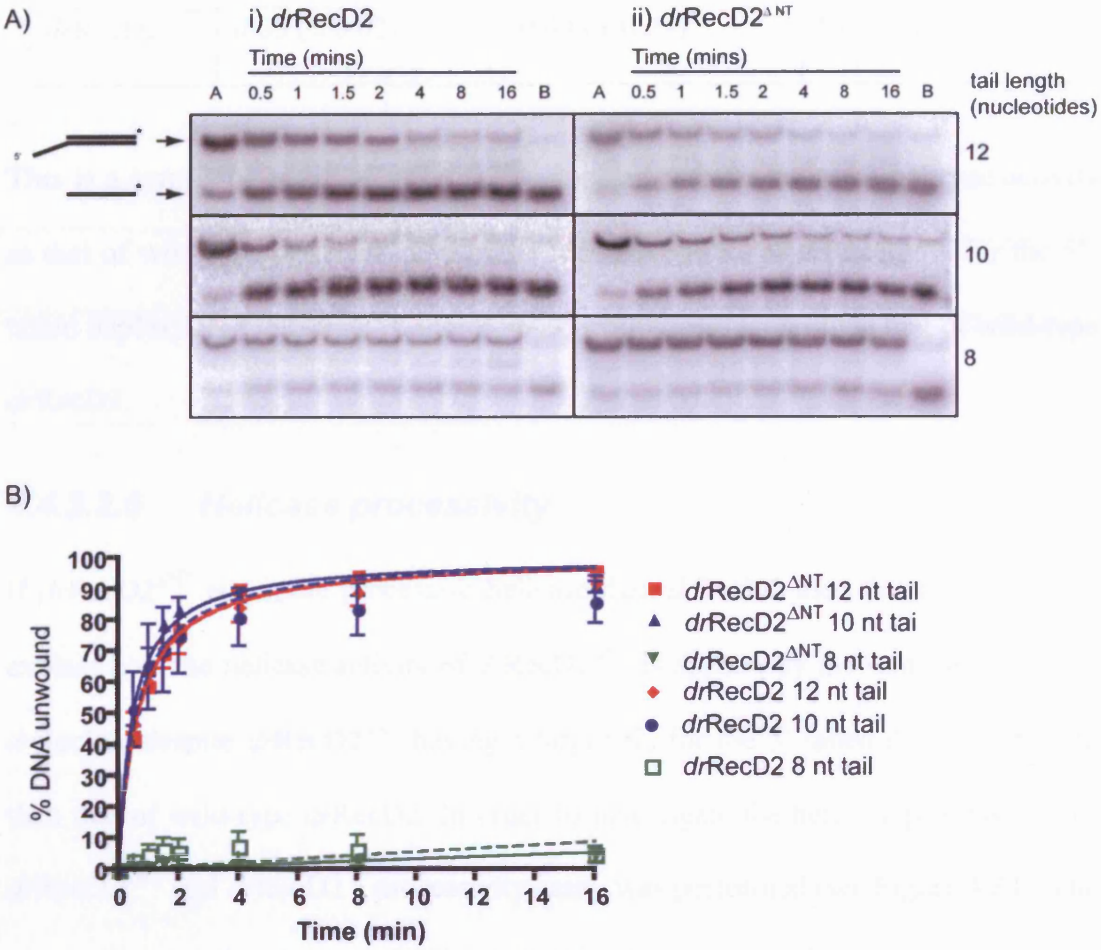

Figure 4.33. Helicase activity of *dr*RecD2$^{\Delta NT}$

A) Autoradiographs of 10% SDS-PAGE analysis of the helicases assays of i) *dr*RecD2 and ii) *dr*RecD2$^{\Delta NT}$. The substrates were all 5'-tailed duplex DNA, the length of the 5' tail varied from 8 to 12 nucleotide (nt). 'A' indicates annealed substrate at time 0 min. 'B' indicates boiled substrate as marker for ssDNA. * indicates radio-labelled DNA.

B) A plot of % DNA unwound against time.

There was no significant difference of the $t_{1/2}$ of *dr*RecD2$^{\Delta NT}$ compared to wild-type *dr*RecD2 (see Table 4.10). Neither was there a difference in the length of 5' tail required for activity.

Table 4.10. Summary of helicase activity of *dr*RecD2$^{\Delta NT}$

	$t_{1/2}$ for pdT12 5'-tailed duplex DNA (min)	**$t_{1/2}$ for pdT10 5'-tailed duplex DNA (min)**	**$t_{1/2}$ for pdT8 5'-tailed duplex DNA (min)**
*dr*RecD2	0.66 (± 0.04)	0.65 (± 0.13)	No activity
*dr*RecD2$^{\Delta NT}$	0.65 (± 0.02)	0.49 (± 0.03)	No activity

This is a surprising result as *dr*RecD2$^{\Delta NT}$ appears to have the same helicase activity as that of wild-type *dr*RecD2 despite the fact that the K_d of *dr*RecD2$^{\Delta NT}$ for the 5'-tailed duplex DNA substrate is approximately 500 times larger than that of wild-type *dr*RecD2.

4.4.3.2.5 Helicase processivity

If *dr*RecD2$^{\Delta NT}$ is a more processive helicase than *dr*RecD2 then this fact may help explain how the helicase activity of *dr*RecD2$^{\Delta NT}$ is apparently the same as wild-type *dr*RecD2, despite *dr*RecD2$^{\Delta NT}$ having a larger K_d for the 5'-tailed duplex substrate than that of wild-type *dr*RecD2. In order to investigate the helicase processivity of *dr*RecD2$^{\Delta NT}$ and *dr*RecD2 a processivity assay was performed (see Figure 4.34). The assay measures the amount of DNA substrate unwound by a helicase, which is pre-bound to the DNA substrate, before the helicase dissociates from the DNA substrate. After dissociating from the DNA the helicase is bound by a trap that stops the re-binding of the helicase to the DNA substrate.

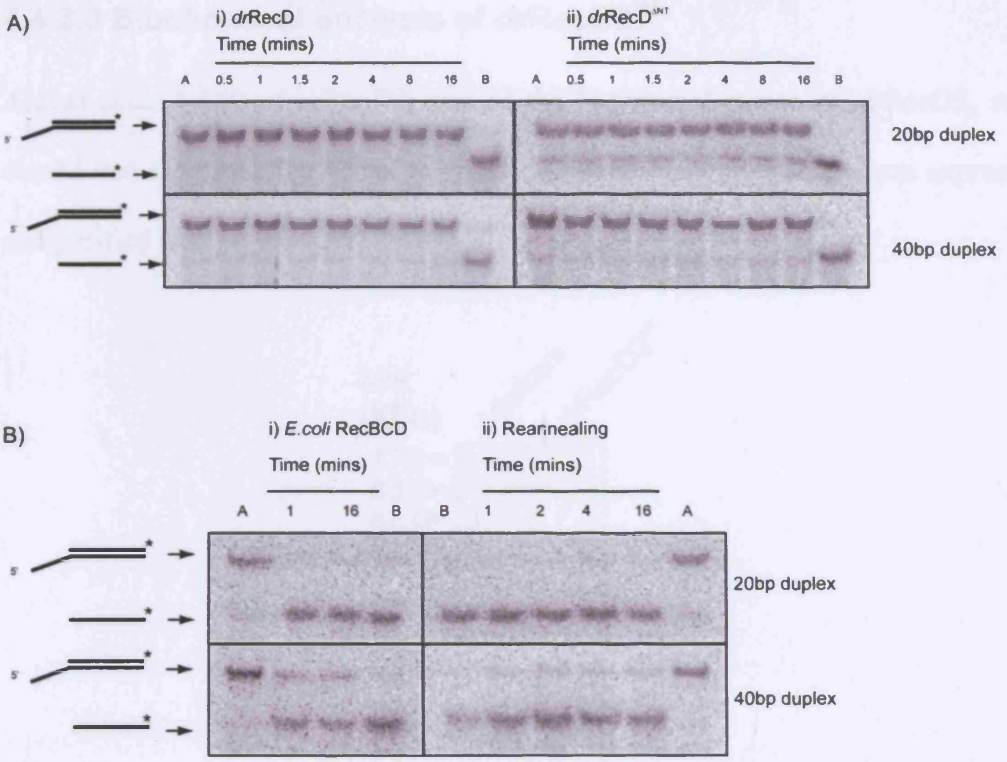

Figure 4.34. Processivity assay of *dr*RecD2$^{\Delta NT}$ and *dr*RecD2

Autoradiograph of 10% SDS-PAGE analysis of :

A) Processivity assays of i) *dr*RecD2 and ii) *dr*RecD2$^{\Delta NT}$ with a 20bp or 40bp 5'-tailed duplex DNA substrate.

B) i) Positive control showing *E.coli* RecBCD1 unwinding both substrates used in the assay. ii) Negative control showing no re-annealing of the DNA substrate during the time course of the assay after boiling. 'A' indicates annealed substrate at time 0 min. 'B' indicates boiled substrate as marker for ssDNA. *indicates radio-labelled DNA.

The assay is dependent on all of the DNA molecules present in the assay being bound to the helicase. In order to achieve this, 1 μM *dr*RecD2$^{\Delta NT}$ was used meaning all the 5'-tailed duplex substrate should be bound (see Figure 4.31). The same concentrations were used for *dr*RecD2 so that a direct comparison could be made. Neither *dr*RecD2$^{\Delta NT}$ nor *dr*RecD2 could unwind 20bp before dissociating from the DNA and no difference could be seen between them. *E.coli* RecBCD1 is a very processive helicase and was therefore used as the positive control.

4.4.3.3 Biochemical analysis of *dr*RecD2^{NT}

Amino acids 1-150 of *dr*RecD2, part of the N-terminal region of *dr*RecD2, were cloned into an expression vector to create *dr*RecD2^{NT} and the protein was expressed and purified (see Figure 4.35).

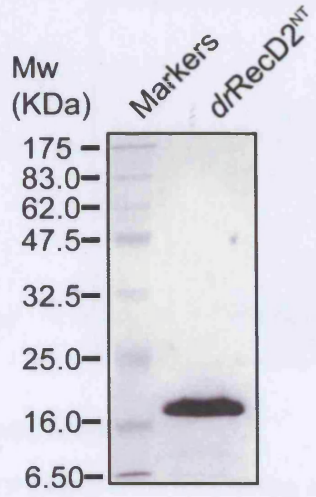

Figure 4.35. Purity of *dr*RecD2^{NT}

A 12% SDS-PAGE analysis of the purity of *dr*RecD2^{NT}.

There were no grounds to suspect that the *dr*RecD2^{NT} would have ATP hydrolysis activity as *dr*RecD2^{NT} does not posses the Walker A and B motifs characteristic of such activity, nor were there any grounds to suspect helicase activity.

DNA binding assays were carried out to see if *dr*RecD2^{NT} could bind to any of the DNA substrates as suggested by the fact that *dr*RecD2$^{\Delta NT}$, which is missing amino acids 1-150, has significant defects in DNA binding compared to wild-type *dr*RecD2. The results show that *dr*RecD2^{NT} failed to bind to any of the DNA substrates (see Figure 4.36), but there is no positive control to confirm that *dr*RecD2^{NT} is correctly folded. No ITC analysis was performed because *dr*RecD2^{NT} had failed to show any DNA binding properties.

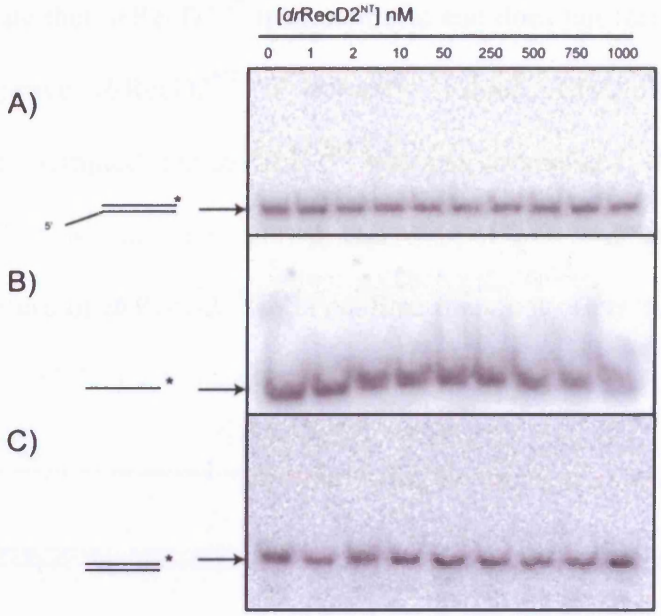

Figure 4.36. DNA binding activity of *dr*RecD2^{NT}

Autoradiographs of 9% native PAGE analysis of the DNA binding assays of *dr*RecD2^{NT} with A) the 5'-tailed duplex. B) the ssDNA and C) the dsDNA substrates. *indicates radio-labelled DNA substrate.

During the purification of *dr*RecD2^{NT} a gel-filtration chromatography step was used. *dr*RecD2^{NT} eluted at approximately 16 ml indicating a molecular weight of approximately 20 KDa suggesting that *dr*RecD2^{NT} is running on the gel-filtration column as a monomer (see Figure 4.37).

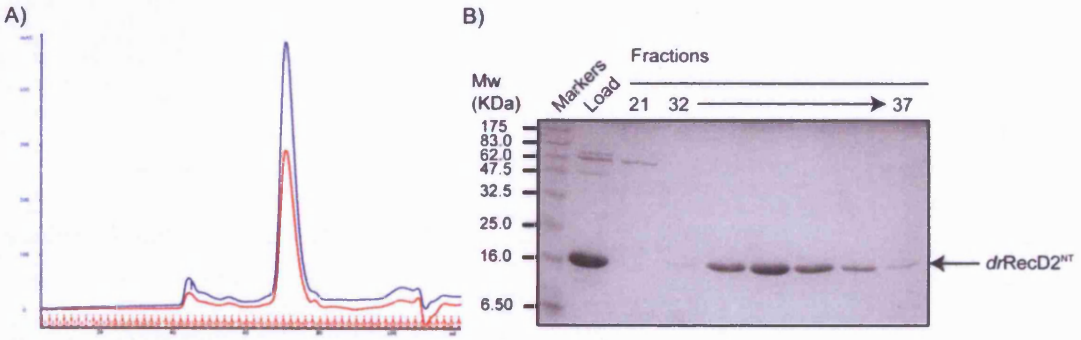

Figure 4.37. Gel-filtration analysis of *dr*RecD2^{NT}

A) The A_{280} (blue) and A_{260} (red) trace of the gel-filtration chromatography of *dr*RecD2^{NT}.

B) A 15% SDS-PAGE analysis of the gel-filtration chromatography showing the load on to the gel filtration column and the significant fractions eluted from the column.

These data indicate that *dr*RecD2^{NT} is monomeric and does not form aggregates, but this does not prove *dr*RecD2^{NT} is correctly folded. Crystallisation trials of *dr*RecD2^{NT} were attempted, but *dr*RecD2^{NT} was not crystallised. If *dr*RecD2^{NT} had been crystallised it would have shown that *dr*RecD2^{NT} was uniformly folded. However, the failure of *dr*RecD2^{NT} to crystallise does not prove that *dr*RecD2^{NT} is misfolded.

4.4.4 The 1B domain

4.4.4.1 Introduction

Recent work on the *E.coli* RecBCD1 structure revealed a loop that is inserted into domain 1A of *E.coli* RecD1, against which the 5' tail of the ssDNA, fed by *E.coli* RecC to *E.coli* RecD1, appears to slide (personal communication, Wigley. D.). This loop (domain 1B) might act as a molecular pin in the RecD2 family over which the DNA duplex is split as part of the helicase mechanism (see Figure 4.38A).

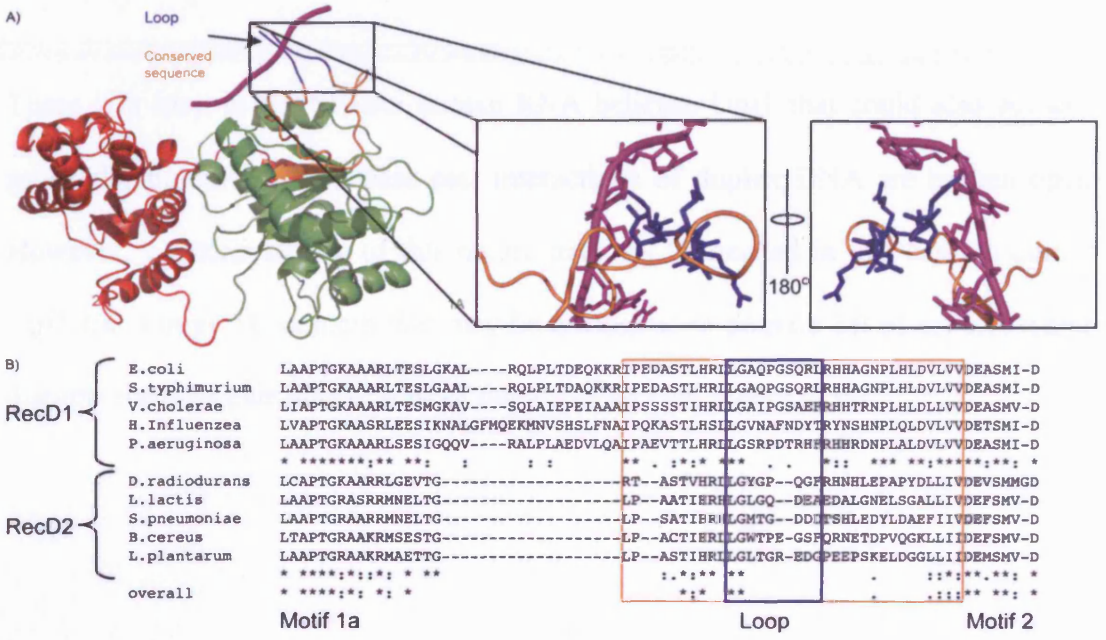

Figure 4.38. The 1B domain loop of RecD

A) The structure of motor domains 1A (green) and 2A (red) of *E.coli* RecD1 with the details of the 1B domain loop (blue) shown. A sequence conserved region is coloured in orange.

B) A sequence comparison of RecD1 and RecD2 between helicase motifs 1a and 2. The domain 1B loop is outlined in blue. The conserved region as shown in A) is outlined in orange.

There does not appear to be any base stacking or hydrogen bond interactions between the amino acids of the loop and the DNA nucleotides, only Van der Waals forces are at play. This suggests that the DNA is not being bound tightly or moved by the loop, but rather it suggests that the loop is a mechanical feature over which

the base pair interactions of duplex DNA are broken apart. The loop could be structurally conserved in other RecD genes and the sequence N-terminal and C-terminal to the loop is conserved in the RecD1 family and *dr*RecD2. However, in those RecD2 helicases that are phylogenetically further away from the RecD1 family (see Chapter 1) only the sequence N-terminal to the loop is conserved (see Figure 4.38B). The RecD1 helicases also have an extra insertion N-terminal to the loop compared to the RecD2 helicases. The loop may, therefore, be in different positions in the RecD1 helicases compared to the RecD2 helicases (see Figure 4.38B).

There is a loop in the SF1Bα human RNA helicase Upf1 that could also act as a molecular pin over which base pair interactions of duplex DNA are broken open. However, a molecular pin of this nature may not be needed in this case as human Upf1 has a large 1C domain that may be analogous to domain 2B of *bs*PcrA which disrupts the base pair interactions of duplex DNA (see Figure 4.39).

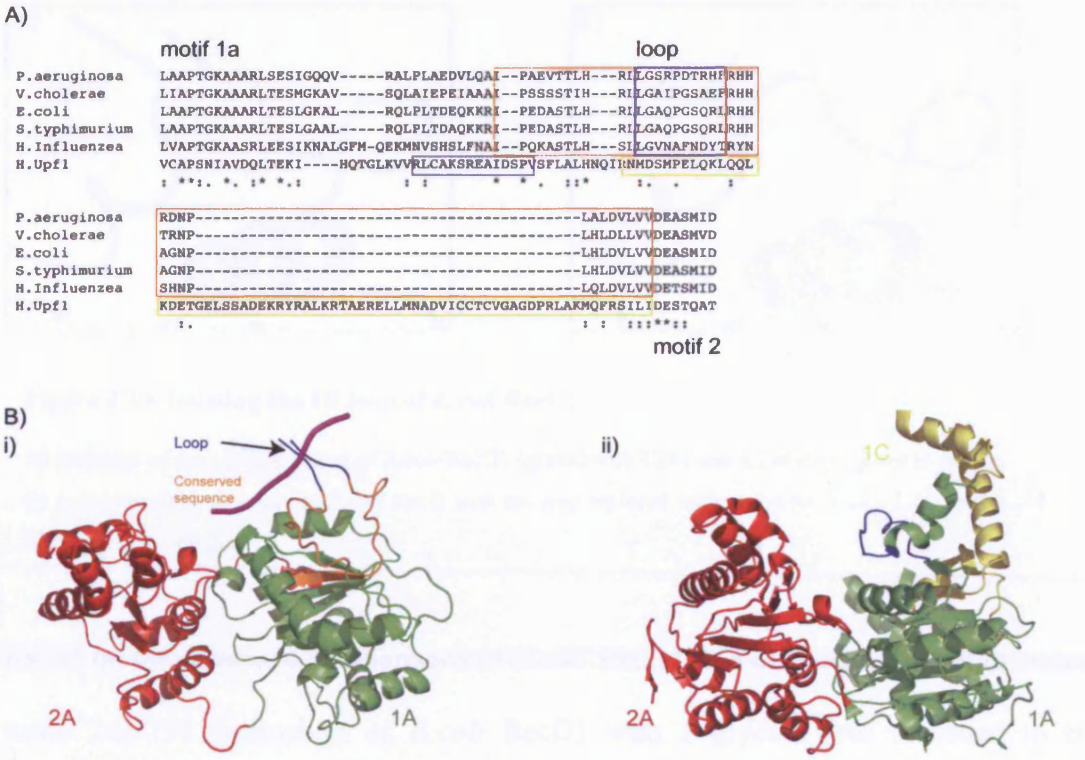

Figure 4.39. A comparison of RecD1 and Upf1

A) A sequence comparison of the RecD1 family and human Upf1. The conserved regions (orange) and loop (blue) of RecD1 are outlined, as is a loop (blue) in Upf1. The 1C domain of human Upf1 is outlined in yellow.

B) A structural comparison of domain 1A and domain 1B of i) *E.coli* RecD1 and ii) Upf1, showing domain 2A (red), 1A (green), 1C (yellow), with the 1B domain loop of *E.coli* RecD1 and a loop in a similar position in Upf1 both coloured blue.

It was reasoned that amino acids 244 to 253 inclusive of *E.coli* RecD1 could be replaced with a glycine in order to remove the DNA interacting part of the loop without disturbing the secondary structure or folding of *E.coli* RecD1. A glycine would be able to span the distance between leucine 243 and arginine 254 in *E.coli* RecD1 (see Figure 4.40).

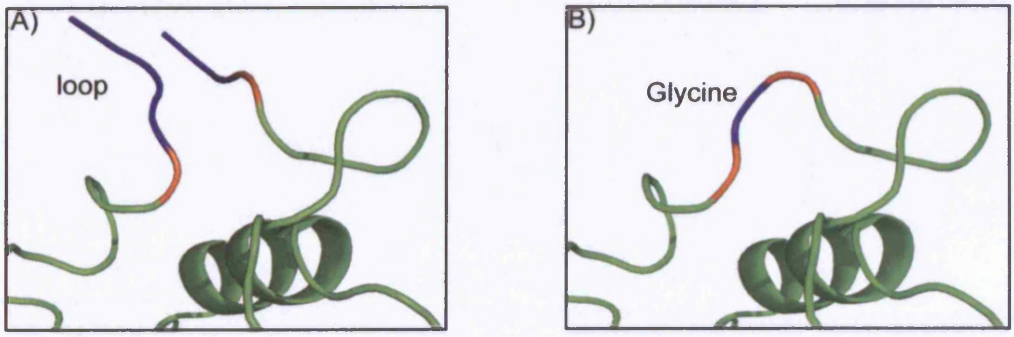

Figure 4.40. Deleting the 1B loop of *E.coli* RecD1

A) Structure of the 1B loop (blue) of *E.coli* RecD1 (green) with L243 and R244 highlighted in orange.
B) A model of the structure of *E.coli* RecD with the loop replaced with a glycine (blue). L243 and R244 highlighted in orange.

Based on the sequence comparisons of *E.coli* RecD1 and *dr*RecD2, replacing amino acids 243-254 (inclusive) of *E.coli* RecD1 with a glycine was reasoned to be analogous to replacing amino acids 412-419 (inclusive) of *dr*RecD2 with a glycine (see Figure 4.38).

4.4.4.2 Biochemical analysis of *dr*RecD2$^{\Delta loop}$

Amino acids 412-419 (inclusive) of *dr*RecD2 were replaced with a glycine to produce *dr*RecD2$^{\Delta loop}$. The protein was expressed and purified (see Figure 4.41).

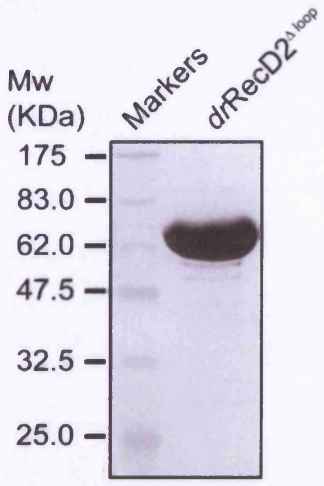

Figure 4.41. Analysis of the purity of *dr*RecD2$^{\Delta loop}$

A 12% SDS-PAGE analysis of the purity of *dr*RecD2$^{\Delta loop}$.

4.4.4.2.1 ATP hydrolysis activity

The ATP hydrolysis activity of *dr*RecD2$^{\Delta loop}$ was investigated and the K_{DNA} and K_{ATP} determined (see Figure 4.42).

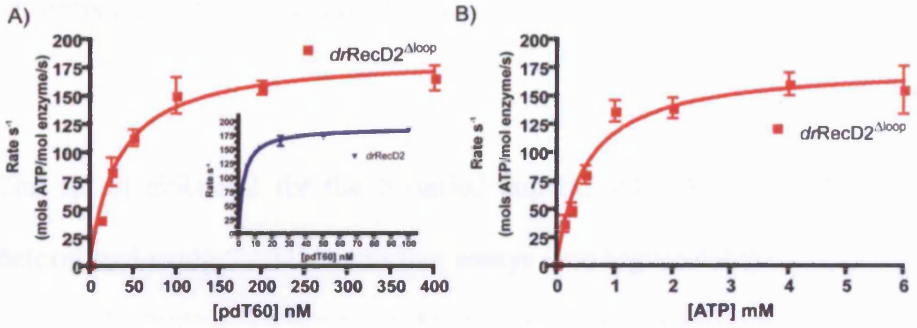

Figure 4.42. ATP hydrolysis activity of *dr*RecD2$^{\Delta loop}$

A) A plot of moles of ATP hydrolysed per mole of *dr*RecD2$^{\Delta loop}$ per second against pdT60 concentration in order to calculate the k_{cat} and K_{DNA} for the reaction in the presence of saturating (4 mM) ATP. The *dr*RecD2 ATP hydrolysis assay is inset for comparison.

B) A plot of moles of ATP hydrolysed per mole of *dr*RecD2$^{\Delta loop}$ per second against ATP concentration in order to calculate the K_{ATP} of the reaction in the presence of saturating (400 nM) pdT60 DNA substrate.

The K_{DNA} of *dr*RecD2$^{\Delta loop}$ was approximately 10 times larger than the K_{DNA} of *dr*RecD2. There was no difference in the k_{cat} and K_{ATP} of *dr*RecD2$^{\Delta loop}$ and *dr*RecD2 (see Table 4.11).

Table 4.11. Summary of ATP hydrolysis activity of *dr*RecD2$^{\Delta loop}$

	K_{DNA} (nM)	k_{cat} (s^{-1})	K_{ATP} (mM)
*dr*RecD2	3.2 (± 0.3)	188 (± 3)	0.57 (±0.03)
*dr*RecD2$^{\Delta loop}$	32 (± 4)	185 (± 5)	0.52 (±0.07)

These data indicate the ATP hydrolysis activity was not affected by the mutation when enough ssDNA was present to overcome the defect in DNA binding and stimulate the ATP hydrolysis activity of *dr*RecD2$^{\Delta loop}$.

4.4.4.2.2 DNA binding activity

The yield of *dr*RecD2$^{\Delta loop}$ from the expression and purification of the protein was approximately 0.1 mgs per litre of *E.coli* culture. This was much lower than the yield of *dr*RecD2 and not enough *dr*RecD2$^{\Delta loop}$ could be purified for an ITC analysis.

The K_d of *dr*RecD2 for the 5'-tailed duplex, ssDNA and dsDNA substrates were determined using the DNA binding assays (see Figure 4.43).

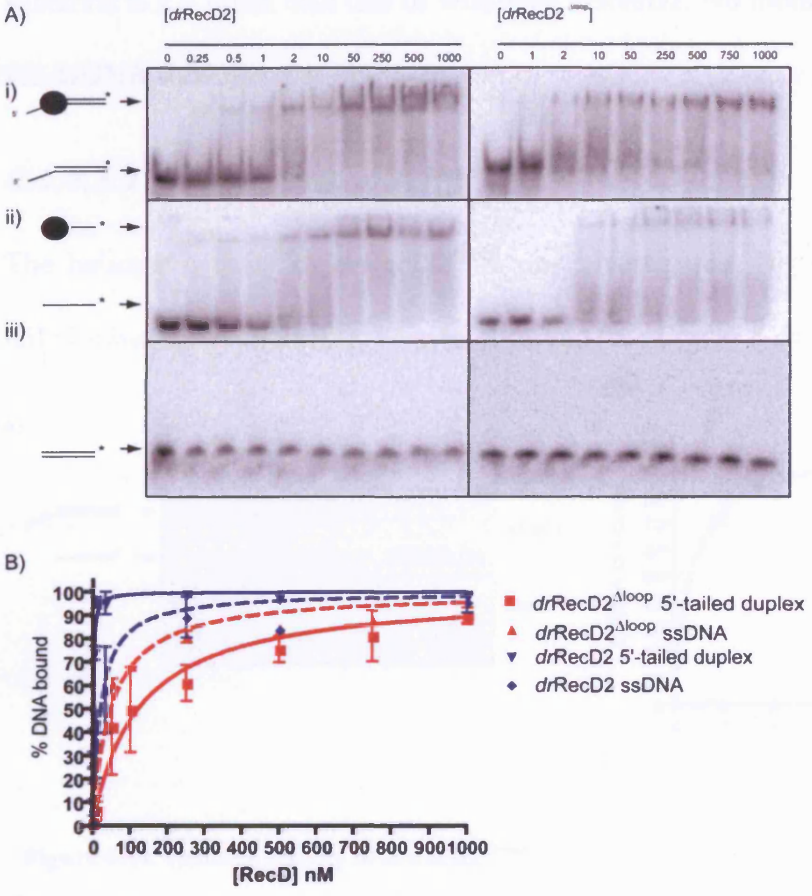

Figure 4.43. DNA binding activity of *dr*RecD2$^{\Delta loop}$

A) Autoradiographs of 9% native PAGE analysis of the DNA binding assays of *dr*RecD2$^{\Delta loop}$ with i) the 5'-tailed duplex, ii) the ssDNA, and iii) the dsDNA substrates. *indicates radio-labelled DNA substrate.
B) Plot of % DNA bound against *dr*RecD2$^{\Delta loop}$ concentration in order to calculate the K_d of *dr*RecD2$^{\Delta loop}$ for the 5'-tailed duplex and ssDNA substrates.

Table 4.12. Summary of the DNA binding activity of *dr*RecD2$^{\Delta loop}$

	K_d for the 5'-tailed duplex substrate (nM)	K_d for the ssDNA substrate (nM)	K_d for the dsDNA substrate (nM)
*dr*RecD2	1.3 (± 0.2)	20 (± 3)	No binding
*dr*RecD2$^{\Delta loop}$	120 (± 12)	48 (± 5)	No binding

The K_d of *dr*RecD2$^{\Delta loop}$ for the 5'-tailed duplex substrate is approximately 100 times larger than that of wild-type *dr*RecD2. The K_d of *dr*RecD2$^{\Delta loop}$ for the ssDNA substrate is 2.5 times than that of wild-type *dr*RecD2. No binding of *dr*RecD2$^{\Delta loop}$ to the dsDNA substrate was observed.

4.4.4.2.3 Helicase activity

The helicase activity of *dr*RecD2$^{\Delta loop}$ was investigated. No unwinding of the 5'-tailed substrate by *dr*RecD2$^{\Delta loop}$ was observed (see Figure 4.44).

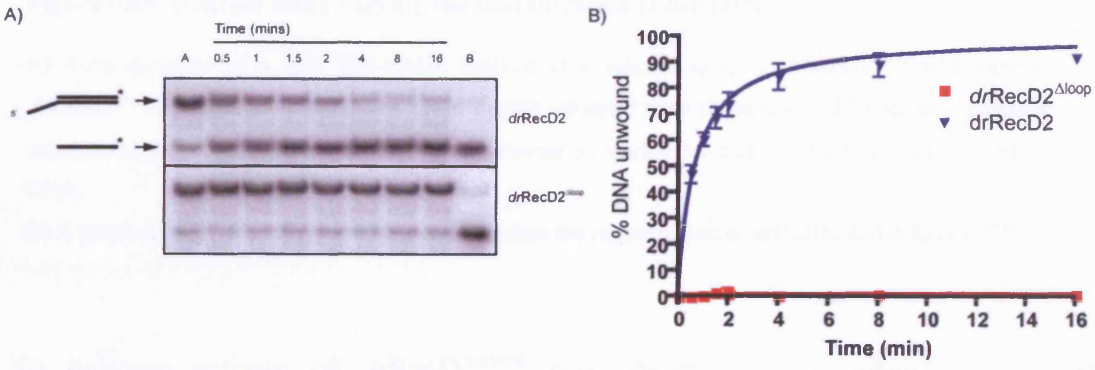

Figure 4.44. Helicase activity of *dr*RecD2$^{\Delta loop}$

A) Autoradiographs of 10% SDS-PAGE analysis of the helicase assays of *dr*RecD2$^{\Delta loop}$ and *dr*RecD2. 'A' indicates annealed substrate at time 0 min. 'B' indicates boiled substrate as marker for ssDNA. *indicates radio-labelled DNA.

B) A plot of % DNA substrate unwound against time.

The lack of helicase activity could have been a symptom of *dr*RecD2$^{\Delta loop}$ not being able to bind to the 5'-tailed duplex DNA substrate. Both the helicase and DNA

binding assays use 1 nM of the 5'-tailed duplex DNA substrate. The DNA binding assays show that when 1 nM *dr*RecD2$^{\Delta loop}$ is used the DNA substrate is not bound by *dr*RecD2$^{\Delta loop}$, but if 1 μM *dr*RecD2$^{\Delta loop}$ is used then the DNA substrate is bound by *dr*RecD2$^{\Delta loop}$ (see Figure 4.43 A). Therefore, a helicase assay was performed in which the concentration of *dr*RecD2$^{\Delta loop}$ was varied and the time was kept constant at 16 minutes (see Figure 4.45).

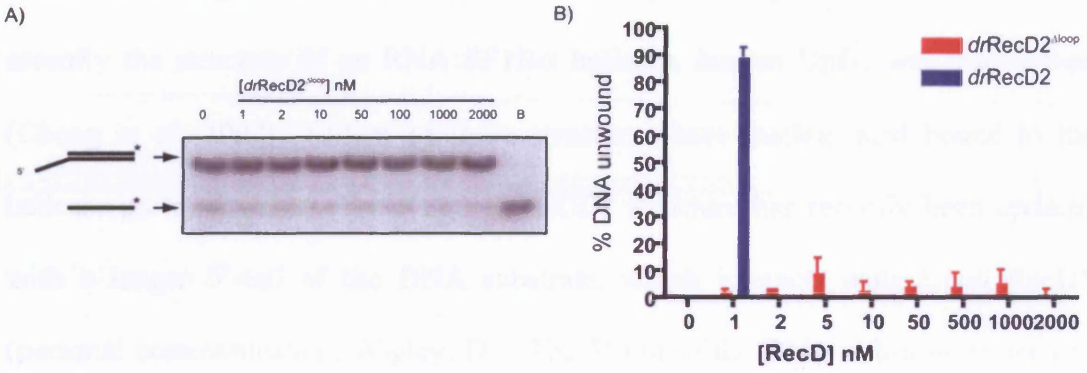

Figure 4.45. Helicase assay varying the concentration of *dr*RecD2$^{\Delta loop}$

A) Autoradiograph of a 10% SDS-PAGE analysis of a helicase assay of *dr*RecD2$^{\Delta loop}$ with varying *dr*RecD2$^{\Delta loop}$ concentration and 1 nM tailed duplex substrate over 16 minutes. 'A' indicates annealed substrate at time 0 min. 'B' indicates boiled substrate as marker for ssDNA. *indicates radio-labelled DNA.

B) A graph of % DNA unwound in 16 minutes against the concentration of *dr*RecD2 and *dr*RecD2$^{\Delta loop}$.

No helicase activity of *dr*RecD2$^{\Delta loop}$ was observed, even when 1 - 2 μM *dr*RecD2$^{\Delta loop}$ was used so that all the tailed duplex substrate present was bound by *dr*RecD2$^{\Delta loop}$.

These data suggest that the defect in the helicase activity of *dr*RecD2$^{\Delta loop}$ is due to an intrinsic helicase defect and not due to ATP hydrolysis or DNA binding defects.

4.5 Discussion

The aim of this part of the thesis was to gain insights into the helicase mechanism of *Deinococcus radiodurans* RecD2 (*dr*RecD2), a Superfamily1Bα helicase, using biochemical techniques. There is very little known about the SF1Bα helicase mechanism and only one structure of a DNA SF1Bα helicase is available, that of *E.coli*RecD1 as part of the *E.coli* RecBCD1 complex (Singleton *et al.* 2004). More recently the structure of an RNA SF1Bα helicase, human Upf1, was also solved (Cheng *et al.* 2007). Neither of these structures have nucleic acid bound to the helicase motor domains. The *E.coli* RecBCD1 structure has recently been updated with a longer 5'-tail of the DNA substrate, which interacts with *E.coli* RecD1 (personal communication, Wigley, D.). The 5' tail of the DNA substrate is seen to interact with domain 1A of *E.coli* RecD1, but does not reach the domain 2A of the helicase, and the helicase mechanism cannot be elucidated.

4.5.1 Identification and purification of *Deinococcus radiodurans* RecD2

Deinococcus radiodurans RecD2 was chosen as the target for study as it was the only soluble RecD helicase identified in this thesis. It was decided to screen RecD2 helicases because they might be more soluble than RecD1 helicases. The reasoning was that RecD1 enzymes exist as part of a RecBCD1 complex and would be insoluble when separated from the complex, and because RecD2 proteins are not part of a RecBCD1 complex they may be soluble. However, all but one of the RecD2 helicases cloned in this work were insoluble. Predicting whether a protein will be soluble is not an exact science and the reasons for the majority of the RecD2

proteins tested in this work being insoluble are not clear. It is tempting to speculate that RecD2 helicases do exist as part of a larger order complex in the cell.

The classification of RecD-like enzymes may need to be further developed. It seems illogical to refer to RecD2 enzymes as RecD because there is no evidence to suggest that RecD2 enzymes perform an analogous function to RecD1 enzymes. There may be more RecD-like enzymes that do fit into the RecD1 or RecD2 family, for example, the YF19_METJA protein of the archaeon *Methanococcus jannaschii*. Eukaryotes and archaea do not use the RecBCD1 reaction pathway so even if RecD2 helicases performed an analogous function to RecD1 helicases, this cannot be the case for the *Methanococcus jannaschii* YF19_METJA protein.

The purification of *dr*RecD2 was based upon work by Wang *et al.* (Wang and Julin 2004). It was apparent that the purification described by Wang *et al.* was not designed to produce the high yields of protein required for crystallography. The first improvement on the protocol was the use of the *E.coli* BL21 Rosetta strain to express the protein, which yielded more *dr*RecD2 protein than using BL21. The second improvement was the use of a heparin column rather than a ssDNA cellulose column. The ssDNA cellulose is a single use medium and would need to be replaced for every purification. Apart from being expensive, this does not lend itself to a high throughput purification protocol because the preparation of the medium is time consuming and the ssDNA can be stripped off the cellulose very easily. A heparin column served the same purpose and was readily available. The third improvement was to add a MonoQ column to the protocol, which further purified *dr*RecD2 to the extent required to start crystallisation trials.

Two interesting observations were made from the purification of *dr*RecD2. Firstly, a fragment of *dr*RecD2 was identified as being an N-terminal truncation of *dr*RecD2. The N-terminal truncated *dr*RecD2 fragment was N-terminally sequenced. A consistent and well defined N-terminal sequence was identified suggesting the majority of the *dr*RecD2 molecules were proteolysed at the same point between amino acids 150 and 151. This is a very specific proteolysis and suggested that amino acids 1-150 form a region of secondary structure that is linked to the rest of the protein with a flexible linker that is open to proteolysis. The gene sequence of the *dr*RecD2 fragment was cloned into an expression vector to create *dr*RecD2$^{\Delta NT}$. The second observation was a possible interaction of *dr*RecD2 with *E.coli* RecB and RecC, suggested by the fact that they all co-purified on the MonoQ column in a separate peak from the main peak of *dr*RecD2 (see Figure 4.6 B). However, this interaction could not be reproduced using gel filtration analysis. This does not discount an interaction of *dr*RecD2 and *E.coli* RecB and C, indeed *dr*RecD2$^{\Delta NT}$ was also seen to co-purify with *E.coli* RecB and RecC on a MonoQ column. There could be *in vivo* factors within the *E.coli* expression strain that are aiding the complex formation of *dr*RecD2 and *E.coli* RecBC , which are missing in the *in vitro* experiments

4.5.2 Biochemical analysis of wild-type *Deinococcus radiodurans* RecD2

Helicase activity is reliant on the co-ordination of several molecular processes. One of these activities is ATP hydrolysis that provides the energy required to unwind duplex DNA and to translocate ssDNA. Another is the DNA binding activity, in order for the helicase to be able to unwind a substrate the helicase must be able to bind to the substrate. In order to understand the effects of any mutations of *dr*RecD2

on the helicase activity, the ATP hydrolysis, DNA binding and helicase activities of wild-type *dr*RecD2 were characterised.

The k_{cat}, K_{ATP} and K_{DNA} were all determined for wild-type *dr*RecD2. Other work has shown *dr*RecD2 to be a ssDNA-stimulated ATPase, but no kinetic analysis was undertaken (Wang and Julin 2004) so these results cannot be compared to the kinetic analysis of this chapter. Many SF1 helicases are ssDNA-dependent ATPases and *dr*RecD2 appears to be a better ssDNA stimulated ATPase than *bs*PcrA. The k_{cat} of *dr*RecD2 is around two orders of magnitude higher than that of *bs*PcrA under ATP and DNA saturating conditions (Dillingham *et al.* 1999), although differences in the buffer conditions have not been controlled for.

In the DNA binding assays the K_d is taken as equal to the *dr*RecD2 concentration required to bind 50% of the total DNA substrate to *dr*RecD2. This is not ideal and will not provide as accurate a K_d as isothermal calorimetry. This is because the assumption has to be made that the free DNA and protein/DNA complex is in equilibrium. However this equilibrium will be disturbed as the protein/DNA complex is separated away from the free DNA and free protein in the electrophoresis gel. However, the assay is very useful to compare the relative DNA binding activity of different proteins under identical conditions. These assays were ideally suited to comparing the wild-type *dr*RecD2 to the *dr*RecD2 mutants.

As expected the DNA binding assays show that wild-type *dr*RecD2 binds to the ssDNA and the 5'-tailed duplex DNA substrates. Previous work by Wang *et al.* had used hairpin substrates to create pseudo 5'-tailed duplex DNA substrates (Wang and Julin 2004), but no K_d of *dr*RecD2 for the DNA substrates is presented for

comparison with those of this chapter. Details of ssDNA and dsDNA binding by *dr*RecD2 have not previously been reported. This chapter shows that *dr*RecD2 prefers to bind to the 5'-tailed duplex substrate than to the ssDNA substrate and cannot bind dsDNA at all. Therefore, *dr*RecD2 must be binding the structure of the tailed duplex junction between ssDNA and dsDNA with a higher affinity than simply the sum of the dsDNA and ssDNA binding affinities. It is highly likely that *dr*RecD2 is unwinding base pairs at the junction of ssDNA and dsDNA. It is common for helicases to unwind a few base pairs of the substrate upon binding without the need for ATP hydrolysis (Singleton *et al.* 2004; Singleton *et al.* 2001; Velankar *et al.* 1999). The ITC data shows that as the length of ssDNA increases from 8 to 15 nucleotides *dr*RecD binds the substrate with higher affinity. The optimum length appears to be around 15 nucleotides indicating that this is sufficient for all the DNA-binding pockets of *dr*RecD2 to be bound to DNA. This is consistent with *dr*RecD2 unwinding 3 base pairs of the 5'-tailed duplex DNA substrate upon binding, increasing the length of the ssDNA tail from 12 to 15 nucleotides.

The ITC experiments show that 1 molecule of *dr*RecD2 can bind to 1 molecule of pdT 10, pdT 12 or pdT15 and that 2 molecules of *dr*RecD2 will bind to one molecule of pdT25 or pdT30 with out any co-operative binding. The ITC data also shows that two binding events occur when *dr*RecD2 binding to pdT20, pdT21, pdT22, pdT 23 or pdT 24 is analysed. These two binding events fit with the theory that if *dr*RecD2 is in excess of the DNA substrate, two molecules of *dr*RecD2 can bind to one molecule of DNA with the equilibrium constant K_{d1}. Later in the ITC experiment when DNA is in excess of *dr*RecD2, one of the two *dr*RecD2 molecules bound to one DNA molecule dissociates and binds to a free DNA molecule with the equilibrium constant K_{d2}. The errors are very large when looking at the two events,

because two curves are being fitted using the same amount of data as fitting one curve in other experiments. Despite the large errors the results show that although two molecules *dr*RecD2 can bind to one DNA molecule when *dr*RecD2 is in excess, one molecule of *dr*RecD2 still prefers to bind per DNA molecule when the DNA is in excess. Once the DNA length increases from 20 to 25 (or more) nucleotides, two molecules of *dr*RecD2 will bind per DNA and only one binding event is observed. There is, therefore, a transition between binding one molecule of *dr*RecD2 per DNA molecule to binding two molecules of RecD per DNA molecule when the length of DNA is increased from 20 to 25 nucleotides depending on the concentrations of *dr*RecD2 and DNA. These data also suggest that the binding of one *dr*RecD2 molecule a pdT25 or pdT30 molecule does not aid nor hinder the binding of a second *dr*RecD2 molecule to the same DNA molecule, because a two-event-binding model cannot be fitted to the data. In the event of either co-operative or non co-operative binding a two-site-binding model could be fitted because the first binding event would necessarily have a different equilibrium constant, $K_{d1,}$ to the second binding event K_{d2}.

Previous work has shown *E.coli* RecD1 and *dr*RecD2 to be 5' – 3' helicases (Dillingham *et al.* 2003; Wang and Julin 2004). This chapter did not characterise the helicase activity of wild-type *dr*RecD2 in detail, only for comparison with *dr*RecD2 mutants. The activity described in this thesis appears to be lower than the activity of *dr*RecD2 in work done by Wang *et al.* (Wang and Julin 2004), despite the fact that this chapter used the same conditions and substrate. However, it is hard to compare the results of this chapter with the previous work because no initial rate of unwinding or $t_{1/2}$ is presented by Wang *et al.* Wang *et al.* show that 0.25 nM *dr*RecD2 will unwind 80% of the DNA substrate in 2 minutes, but the reaction is not

shown to go to completion. In this chapter 1 nM *dr*RecD2 requires approximately 2.5 minutes to unwind 80% of the DNA substrate before the reaction reaches completion suggesting that the *dr*RecD2 in this chapter is approximately 5 -10 times less active than the *dr*RecD2 previously described, and the difference in helicase activity was a concern. The only difference between this chapter and previous work was the purification of *dr*RecD2. In order to control for this *dr*RecD2 was purified as described by Wang *et al.* (Wang and Julin 2004) and the helicase activity measured (data not shown). The helicase activity was the same as previously described in this chapter. It was therefore reasoned that this chapter would be consistent within itself and the activity of *dr*RecD2 mutants could be compared to the wild-type *dr*RecD2 activity described in this chapter.

A major problem came from the fact that the *dr*RecD2 mutants could not be cloned using the wild-type *dr*RecD2 gene as a template. The cloning of the *dr*RecD2 gene into an expression vector from *Deinococcus radiodurans* genomic DNA proved to be difficult because of the failure of the standard PCR of the gene. The standard PCR of the *dr*RecD2 gene required to make mutants of *dr*RecD2 failed as did the modified PCR used in cloning of wild-type *dr*RecD2. The reason for the failure of the PCR was attributed to the high GC content of the wild-type *dr*RecD2 gene, which is around 70 % GC . For this reason a gene was synthesised that was less GC rich. The *dr*RecD2 protein expressed from the synthetic gene had the same activity as the *dr*RecD2 previously described.

4.5.3 The role of phenylalanine 635 of motif A

This chapter described the effects of mutating phenylalanine 635 in a newly identified motif A that exists in both the RecD1 and RecD2 families. The region in

which the motif is located is disordered and cannot be seen in the crystal structure of the *E.coli* RecBCD1 complex (Singleton *et al.* 2004). The reason phenylalanine 635 was targeted as a priority was because the key phenylalanine 64 of *bs*PcrA motif 1a is missing in the RecD1 and RecD2 helicases. It was reasoned that phenylalanine 635 of motif A may be interacting with ssDNA bases and contributing toward the helicase mechanism. However, when phenylalanine 635 was mutated to an alanine in *dr*RecD2^{F635A} there was no observable difference in ssDNA-dependent ATP hydrolysis or helicase activity as compared to wild-type *dr*RecD2. This would suggest that phenylalanine 635 is not playing a role in helicase activity or that the experiments used are not sensitive enough to detect any defect. Either way the mutation is not having a major effect on the helicase activity of *dr*RecD2. This does not discount motif A from being involved in the helicase activity, but other roles of motif A need to be investigated.

4.5.4 The role of the N-terminal region

The N-terminal region of type 2 RecD helicases is longer than that of the N-terminal domain 1 of type 1RecD helicases. Domain 1 of *E.coli* RecD1 is responsible for binding to *E.coli* RecC in the RecBCD1 complex (Singleton *et al.* 2004). This however cannot be the role of the N-terminal region of the RecD2 helicases because, by definition, there is no RecC to bind to. The aim of this chapter was to gain some insight into the role of the N-terminal region of *dr*RecD2 by making a truncated *dr*RecD2, missing part of the N-terminus, *dr*RecD2$^{\Delta NT}$.

The K_{DNA} of *dr*RecD2$^{\Delta NT}$ was two orders of magnitude greater than the K_{DNA} of wild-type *dr*RecD2 suggesting that *dr*RecD2$^{\Delta NT}$ would have a higher K_d for ssDNA than *dr*RecD2. However, the k_{cat} was unaffected meaning that the ATP hydrolysis

activity was not impaired once enough ssDNA was present to overcome any ssDNA binding defects and saturate the protein. The K_{ATP} was also unaffected, meaning the binding of ATP to the protein was not affected by the loss of the N-terminal domain. These data suggest that the N-terminus of *dr*RecD2 plays no role in the ATP hydrolysis activity of *dr*RecD2.

The K_d of *dr*RecD2$^{\Delta NT}$ for the 5'-tailed duplex substrate was two orders of magnitude larger than that of *dr*RecD2. The K_d of *dr*RecD2$^{\Delta NT}$ for the ssDNA substrate was shown to be at least two orders of magnitude larger than that of *dr*RecD2. These data suggest the N-terminus is playing a role in binding the 5'-tailed duplex junction of dsDNA and ssDNA. It is surprising, therefore, that the helicase activity appears to be unaffected by the truncation, as *dr*RecD2$^{\Delta NT}$ appears to have the same helicase activity as *dr*RecD2. It would follow that if *dr*RecD2$^{\Delta NT}$ could not bind the substrate as well as *dr*RecD2, the helicase activity of *dr*RecD2$^{\Delta NT}$ would be lower than that of *dr*RecD2, but this is not the case. The helicase activity of *dr*RecD2$^{\Delta NT}$ appears to be the same as that of wild-type *dr*RecD2 despite the fact that the K_d of *dr*RecD2$^{\Delta NT}$ for the 5'-tailed duplex DNA substrate is approximately 500 times greater than that of wild-type *dr*RecD2. The only difference in the buffers used for the DNA binding assays and the helicase assays was the addition of ATP in the helicase assay. The ATP is probably not affecting the DNA binding activity of *dr*RecD2$^{\Delta NT}$ as the ATP hydrolysis assays also show a defect in DNA binding, the K_{DNA} of *dr*RecD2$^{\Delta NT}$ was two orders of magnitude larger than that of *dr*RecD2. Another explanation for the increase of helicase activity of *dr*RecD2$^{\Delta NT}$ could be that more *dr*RecD2$^{\Delta NT}$ molecules are bound per DNA substrate. However, the ITC data shows that only one molecule of *dr*RecD2$^{\Delta NT}$ binds per pdT15 molecule, the same as

*dr*RecD2. It is hard to see how more molecules of *dr*RecD2$^{\Delta NT}$ rather than *dr*RecD2 could bind to the 12 nucleotide long ssDNA portion of the 5'-tailed duplex substrate.

Another explanation for *dr*RecD2$^{\Delta NT}$ having the same helicase activity as wild-type *dr*RecD2 could be that *dr*RecD2$^{\Delta NT}$ is a more processive helicase than *dr*RecD2. This would mean that despite the fact *dr*RecD2$^{\Delta NT}$ cannot bind the DNA substrate as well initially, once the helicase activity starts *dr*RecD2$^{\Delta NT}$ may stay bound and unwind more base pairs than *dr*RecD2 can, before falling off. However, there appeared to be no obvious difference between the processivity of *dr*RecD2$^{\Delta NT}$ and *dr*RecD2. In hindsight this may have been expected as the K_{DNA} of ATP hydrolysis is an indirect measure of ssDNA binding activity during ssDNA translocation by the helicase and the K_{DNA} of *dr*RecD2$^{\Delta NT}$ was around two orders of magnitude larger than that of *dr*RecD2.

The N-terminus of *dr*RecD2 may be acting as a regulatory domain that inhibits the helicase of the wild-type protein. When the N-terminal is removed the inhibition is lost, but the DNA binding activity is also reduced so the helicase activity of *dr*RecD2$^{\Delta NT}$ observed using the helicase assay appears to be same as wild-type *dr*RecD2. A gain-of-function mutation has been observed in the Superfamily 1Aα helicase, *E.coli* Rep, where the 2B domain has been shown not to be required for the helicase activity of Rep and to inhibit the helicase activity of Rep when present. (Brendza *et al.* 2005; Cheng *et al.* 2002). Brendza *et al.* speculate that the 2B domain of Rep may be mutually sequestered when Rep forms a dimer negating the inhibition and activating the helicase. It is interesting to speculate that a similar regulatory function may be occurring in the case of *dr*RecD with a speculative binding partner

sequestering the N-terminal region of *dr*RecD2 and also tethering *dr*RecD2 to the DNA.

The aim of another part of this thesis (Chapter 3) was to investigate further the role of the 2B domain in the PcrA helicase that appears to contribute to the helicase activity and not inhibit it. (Ha *et al.* 2002; Soultanas *et al.* 2000; Velankar *et al.* 1999).

In order to investigate further the role of the N-terminal region of *dr*RecD2 in binding DNA amino acids 1-150 of *dr*RecD2, deleted in *dr*RecD2$^{\Delta NT}$, were cloned separately to create *dr*RecD2^{NT}. *dr*RecD2^{NT} did not appear to bind any of the DNA substrates, but there is no positive control available to confirm that *dr*RecD2^{NT} is folded correctly, and so it is hard to draw conclusions from these data.

4.5.5 The role of the 1B domain

The final part of this chapter was to investigate the role of a loop that is probably structurally conserved in the RecD1 and RecD2 families. The sequence of the loop itself appears to vary, but the sequence N-terminal and C-terminal to the loop appear to be conserved. The *E.coli* RecBCD1 structure had has recently been improved with a longer 5'-tailed duplex present in the crystals. This structure shows the ssDNA fed to RecD1 from RecC sliding against a loop in RecD1. It was reasoned that this loop could act as a mechanical pin over which the RecD2 family can force open the duplex DNA, much in the same way as a pin present in *E.coli* RecC (Singleton *et al.* 2004). Indeed this may be how *E.coli* RecD1 is able to unwind a 5'-tailed duplex substrate in isolation (Dillingham *et al.* 2003) without *E.coli* RecC present to split the duplex DNA ahead of *E.coli* RecD1 (Singleton *et al.* 2004). Based on the

structure of *E.coli* RecD1 and sequence comparisons of the RecD1 and RecD2 families, amino acids 412 to 419 of *dr*RecD2 were replaced with a glycine in order to investigate the role of the loop by removing it.

The K_{DNA} of *dr*RecD2$^{\Delta loop}$ was an order of magnitude bigger than the K_{DNA} of wild-type *dr*RecD2 suggesting that *dr*RecD2$^{\Delta loop}$ would have a higher K_d for ssDNA than *dr*RecD2. However, the k_{cat} was unaffected meaning that ATP hydrolysis activity was not impaired once the concentration of ssDNA was high enough to overcome any ssDNA binding defects and saturate *dr*RecD2$^{\Delta loop}$. The K_{ATP} was also unaffected, meaning the binding of ATP to the protein was not affected by the loss of the loop. These data suggest that removal of the loop does not affect the ATP binding or ATP hydrolysis activities of *dr*RecD2.

The K_d of *dr*RecD2$^{\Delta loop}$ for the 5'-tailed duplex substrate was two orders of magnitude larger than that of *dr*RecD2 and the K_d of *dr*RecD2$^{\Delta loop}$ for ssDNA was shown to be twice as large as that of *dr*RecD2. These data show that the ssDNA binding activity is not as affected as the 5'-tailed duplex binding activity by the mutation. As there are only three DNA structures to bind, the ssDNA, the dsDNA and the junction of dsDNA and ssDNA, these data suggest that the loop binds the 5'-tailed duplex substrate at the junction of the dsDNA with ssDNA, because no dsDNA binding was observed. Despite the reduced affinity for the 5'-tailed duplex substrate, if an excess *dr*RecD2$^{\Delta loop}$ (1 μM) was added to 1nM 5'-tailed duplex substrate, then all the 5'-tailed duplex was bound by *dr*RecD2$^{\Delta loop}$.

There was no observable helicase activity by *dr*RecD2$^{\Delta loop}$. When 1μM *dr*RecD2$^{\Delta loop}$ and 1 nM of the 5'-tailed duplex are used, meaning the DNA binding defect has been

overcome and all the DNA is bound by *dr*RecD2$^{\Delta loop}$ and, as already discussed, there is no defect in ATP hydrolysis, the *dr*RecD2$^{\Delta loop}$ still cannot unwind the 5'-tailed substrate. These data show that the loop is essential for helicase activity. The ATP hydrolysis and DNA binding activities of *dr*RecD2 are affected when the loop is removed, however, the defects can be over come when enough DNA substrate is present in the assays. This means that the loss of helicase activity upon removal of the loop from *dr*RecD2 is not simply a result of loss of ATP hydrolysis or DNA binding actives of *dr*RecD2, but rather the intrinsic helicase activity. These data indicate that *dr*RecD2 is an active helicase.

All these data support the hypothesis that the loop is acting as a mechanical pin over which dsDNA is split into separate strands, because the loop is required to bind the ssDNA - dsDNA junction of a tailed duplex substrate and the loop is also essential for helicase activity. However, deleting the loop may have displaced the spatial orientation of the 1a motif, which may play a significant role in ssDNA translocation. Translocation activity is assumed to be unaffected because the ATP hydrolysis activity is only slightly impaired. This assumption may not be true and it is very important that the translocation of ssDNA by *dr*RecD2 be shown experimentally unaffected by the removal of the loop.

4.5.6 Conclusions

The main conclusions of this chapter are: 1) A conserved phenylalanine located in Motif A of *dr*RecD2 does not appear to be required for helicase activity. 2) The N-terminal region of *dr*RecD2 plays a role in binding DNA and may act to inhibit the helicase activity of the protein. 3) A conserved loop in *dr*RecD2 is involved in the

binding of the ssDNA – dsDNA junction of 5' tailed duplex DNA and is essential for helicase activity.

4.6 Future Work

There are experiments that could have been performed to further investigate the helicase mechanism of *dr*RecD2 that were not done because of time constraints.

To further investigate if motif A has any role in the helicase activity other residues could be mutated and the ATP hydrolysis, DNA binding and helicase activities investigated. Motif 3 of *bs*PcrA has a tyrosine, Y257, that is important in creating a ssDNA-binding pocket involved in the translocation of ssDNA by *bs*PcrA. In the RecD1 family the W259 of *bs*PcrA is replaced by a glutamate residue and by an aspartic acid residue in *dr*RecD2, D471. It would be interesting to see if D471 of *dr*RecD2 is involved in the helicase activity of *dr*RecD2 by mutating D471 to an alanine and observing the effects on the ATP hydrolysis, DNA binding and helicase activities of *dr*RecD2

The role of the N-terminal region of *dr*RecD2 can be investigated further firstly by determining a steady-state specific activity of the helicase activity, a measure of how many moles of a DNA substrate can be unwound per mole of *dr*RecD2, or *dr*RecD2$^{\Delta NT}$, per second. This could be done by measuring the initial rate of unwinding (while the substrate is in excess of the helicase) at low concentrations of the helicase compared to the DNA substrate.

An attempt to restore *dr*RecD2$^{\Delta NT}$ DNA binding activity to wild-type *dr*RecD2 levels could be made by trying to complex *dr*RecD2$^{\Delta NT}$ with *dr*RecD2^{NT}. This could

be done by mixing an excess of *dr*RecD2^{NT} to *dr*RecD2$^{\Delta NT}$ under different conditions. Any resulting complex formed could be purified by gel-filtration chromatography. If this was successful DNA binding assays could be carried out to see if DNA binding activity had been reconstituted to wild-type *dr*RecD2 levels. This could be an indication of whether or not *dr*RecD2^{NT} is correctly folded depending on whether the complex is formed and its activity reconstituted.

This chapter only describes one truncation of the N-terminus of *dr*RecD2 and further truncations of the N-terminus could be made until only the helicase domains (domain 1A and domain 2A) of *dr*RecD2 are left. It would be very interesting to see the effect of these mutants on helicase activity.

The conserved 1B domain loop of *dr*RecD2 has been shown to be essential for helicase activity in this chapter. The ATP hydrolysis assay suggests translocation of ssDNA is unaffected, but direct experiments to show this could be done. There are several different ssDNA translocation assays available, the simplest being the observation of streptavidin being displaced from a 3'-biotinylated ssDNA substrate by the translocating enzyme. If the ssDNA-translocation activity of *dr*RecD2 was unaffected by the removal of the loop these data would support the hypothesis that the loop is required to bind and disrupt the dsDNA – ssDNA junction of a 5' tailed duplex substrate and not required for the ssDNA-translocation activity or *dr*RecD2.

In order to further confirm that the loop is required for helicase activity, the loop could be deleted in either or both *E.coli* RecD1 and Upf1 and the ATP hydrolysis, DNA binding and helicase activities measured. In the case of *E.coli* RecD1 this would have to be done with the *E.coli* RecD1 in isolation and not as part of the

complex as in the complex there is no requirement for RecD1 to break open duplex DNA as this is done by *E.coli* RecC. This may prove difficult because the purification of *E.coli*RecD1, in isolation of *E.coli* RecB and *E.coli* RecC, does not yield 100% active protein. Any variations in helicase activity between wild-type *E.coli* RecD1 and mutant *E.coli* RecD1 may be due to the purification and not the mutation. The experiments on *E.coli* RecD1 could be done in the context of the *E.coli* RecBCD1 complex if the helicase activity of RecB is knocked out. An ATP hydrolysis dead *E.coli* RecB mutant is available (Personal communication, Wigley, D.). However, there would be many complications in analysing the effects of any mutations of *E.coli* RecD1 in this case because of the complex interactions between the proteins in the *E.coli* RecBCD1.

Binding partners for *dr*RecD2 could be searched for. This could be done by mixing purified recombinant *dr*RecD2 with *Deinococcus radiodurans* cell extract and then re-purifying *dr*RecD2 which is histidine tagged along with any other proteins bound to *dr*RecD2.

A more ambitious and speculative project could be to investigate the missing phenylalanine of motif 1a of SF1Bα helicases compared to SF1Aα helicases. It would be interesting to see what effect mutating proline 205 of *dr*RecD2 to a phenylalanine would have on helicase activity of *dr*RecD2, and if this could possibly reverse 5' – 3' helicase polarity of *dr*RecD2 to 3' – 5'. However, the mechanism of helicase polarity is probably far more complex than the role of one residue and the 5' – 3' helicase activity of *dr*RecD2 needs to be understood before experiments trying to reverse the helicase polarity are undertaken.

5 Crystallisation of *Deinococcus radiodurans* RecD2

5.1 Introduction

5.1.1 Summary

The aims of this part of the thesis were to crystallise and solve the structure of the RecD helicase in complex with DNA substrates in order to gain insights into the molecular details of the SF1Bα helicase mechanism. The *Deinococcus radiodurans* RecD2 protein was crystallised with a 5'-tailed duplex DNA substrate. Crystals were of the C2 spacegroup with unit cell dimensions a = 169.9 Å, b = 59.6 Å, c = 91.6 Å, α = 90.0 °, β = 98.6 ° γ = 90.0 °, and diffracted to 3.5 Å.

5.1.2 Current structural understanding of Superfamily 1Bα helicases

The only structure of a SF1Bα DNA helicase to date is that of *E.coli* RecD1 as part of the *E.coli* RecBCD1 complex (Singleton *et al.* 2004). More recently the structure of the SF1Bα RNA helicase, Upf1, has been solved in the absence of nucleic acid. (Cheng *et al.* 2007). However, neither of these structures shows nucleic acid bound to the motor domains of the helicases and therefore very little information on the SF1Bα helicase mechanism can be extracted.

5.2 Crystallisation of *Deinococcus radiodurans* RecD2

5.2.1 Original screening

The aim of this work was to crystallise *dr*RecD2 and its DNA complexes. The crystallisation trials required *dr*RecD2 to be concentrated to 10 mg/ml. The concentration had to be done slowly from storage stock of 1 mg/ml over 8-10 hours, as any sudden increase in the protein concentration would lead to the precipitation of *dr*RecD2. The concentration was done using a YM-50 centricon and the protein was buffered in 10 mM Tris HCl pH 7.5 and 100 mM sodium chloride . Crystallisation trials using *dr*RecD2 in the presence and absence of AMPPNP were undertaken. Crystallisation trials of *dr*RecD2 in the presence of ADP were carried out, but *dr*RecD2 precipitated upon addition of ADP. Around a thousand conditions were examined, but no protein crystals were observed.

It is very common for DNA-binding proteins to become more soluble and tractable to work with when bound to a suitable DNA substrate. The DNA substrate used can, therefore, often be the key for growing protein/DNA co-crystals. In order to gain insights into the helicase mechanism it was reasoned that a 5'-tailed duplex substrate would be the best substrate to use because it would show *dr*RecD2 bound to the ssDNA and dsDNA, analogous to the junction between duplex and unwound DNA.

A 5'-tailed duplex DNA substrate was designed (DNA S1) based on the DNA substrate used in the crystal structure of *bs*PcrA. DNA S1 consisted of 10 bp duplex DNA with a pdT7 5' tail (see Figure 5.1).

A)

```
5′ GCAGTGCTCGTTTTTTT 3′
3′ CGTCACGAGC        5′
```

B)

```
5′ TTTTTTTGCAGTGCTCG 3′
3′         CGTCACGAGC 5′
```

Figure 5.1. DNA S1

A) The DNA substrate used in the *bs*PcrA/DNA complex crystals.

B) DNA S1 modified from A) to have a 5', rather than 3', tail.

DNA S1 was suitable for this work for several reasons. Firstly, *dr*RecD2 specifically binds 5'-tailed duplex DNA (see Chapter 4, section 4.3.2.2), and will unwind 5'-tailed duplex DNA in the presence of ATP and magnesium chloride (see Chapter 4, section 4.3.1). Secondly, DNA S1 would show how *dr*RecD2 binds to ssDNA, dsDNA and the ssDNA-dsDNA junction giving insights into the helicase mechanism. The DNA substrate used in the *bs*PcrA crystals formed pseudo homodimers as a result of the crystal packing, with the dsDNA blunt end forming triple-base pair interactions, while *bs*PcrA is bound to the dsDNA – ssDNA junction. The triple-base pair phenomenon could be assisting in the formation of ordered crystal lattices by forming longer regions of duplex DNA and the same could happen to aid *dr*RecD2/DNA S1 crystallisation. The ssDNA tail of 7 nucleotides did not extend out of the *bs*PcrA RecA-like fold core domains and it was reasoned that the same would be the case for *dr*RecD2 because it is very probable *dr*RecD2 will have the same RecA-like fold core domains common to all helicases. If any nucleotides did extend out of the helicase DNA-binding sites they could be disordered and might impair crystal growth.

DNA binding assays were carried out using 1 to 1000 nM *dr*RecD2 and 1 nM DNA S1, but no binding was observed (data not shown). The concentrations at which crystallisation trials were to be carried out (150 to 200 μM *dr*RecD2 and DNA S1) are much higher than 1 μM protein and 1 nM DNA. The binding of *dr*RecD2 to

DNA S1 was therefore tested at higher concentrations using 1 μM DNA and 1 to 10 μM *dr*RecD2. A *dr*RecD2/DNA S1 complex was formed with 100% DNA substrate bound by *dr*RecD2 at stoichiometric concentrations (see Figure 5.2), suggesting the concentrations used for crystallisation trails a *dr*RecD2/DNA S1 complex should form. However, this does not guarantee that under all the crystallisation trial conditions all the *dr*RecD2 would be in complex with DNA S1.

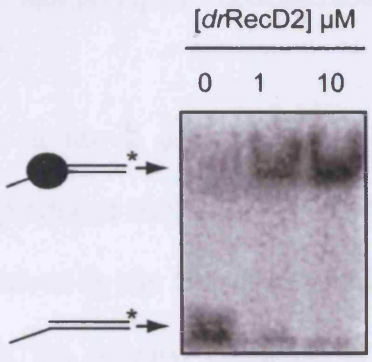

Figure 5.2. Complex formation of *dr*RecD2 and DNA substrate 1

Autoradiograph of a 9% native PAGE analysis of a DNA-binding assay of *dr*RecD2 and DNA S1. 1 μM DNA S1 was used. *indicates radio-labelled DNA.

Crystallisation trials were performed with 150 μM *dr*RecD2 with a 1.3 times molar excess of DNA S1 to try and ensure that all of the *dr*RecD2 was bound to DNA S1. Two crystals were observed in very different conditions. Condition 1 (0.15 M potassium thiocyanate, 0.1 M Tris pH 7.5 and 15% w/v polyethylene glycol 6000 (PEG 6K)) produced clusters of very fine needles. Condition 2 (0.2 M calcium acetate, 0.1 M MES pH 6.5 and 10% w/v PEG 8K) produced clusters of irregularly shaped thin plates with sharp edges. Both of these conditions came from high-throughput screens using the Mosquito robot. Trials were also carried out with the complex in the presence of AMPPNP and magnesium chloride but no crystals were observed. Trials of *dr*RecD2/DNA S1 complex in the presence of ADP were to be carried out, but, as had happened with *dr*RecD2 protein alone, the complex precipitated.

5.2.2 Optimisation

Having identified two crystallisation conditions of the *dr*RecD2/DNA S1 complex in 100 nl sitting drops, attempts to reproduce the crystals in larger 2 µl drops were carried out. The 2 µl hanging drop crystallisation grid screens were carried out based on the conditions identified from the 100 nl sitting drop high-throughput screens. This involved varying the salt and precipitant concentrations of the two conditions.

No crystals were observed in the 2 µl hanging drops based on condition 1. Optimisation of condition 2 produced very clustered needles, but these did not seem to resemble the crystals observed in the 100 nl sitting drops. The crystals grew in various conditions based on the original condition 2. Crystal clusters with the thickest needles grew in 0.1 M MES pH 6.5, 7.5 % PEG 8K and 0.175 M calcium acetate. The crystals typically grew within one week. The crystals were soft and broke easily when probed and were birefringent, suggesting that they were not salt. The clusters did not break into single, useable, crystals when probed, but shattered. The crystals were crudely cryoprotected in 35% glycerol and produced a very low quality diffraction pattern in-house, diffracting to around 8 Å (data not shown).

5.2.3 Growing discrete crystals

The next stage of the optimisation was to grow discrete crystals. The crystallisation process consists of three fundamental stages; nucleation, crystal growth, and cessation of crystal growth. Nucleation is the process by which the first ordered aggregates are formed; these then grow into larger crystals in the crystal growth phase. Nucleation requires greater saturation of the crystallisation target to a thermodynamically unstable state than is required for crystal growth. This means that there is a balance between slowing down crystal growth before stopping

nucleation, and it may be the case that nucleation will be abolished before crystal growth is effected (see Figure 5.3).

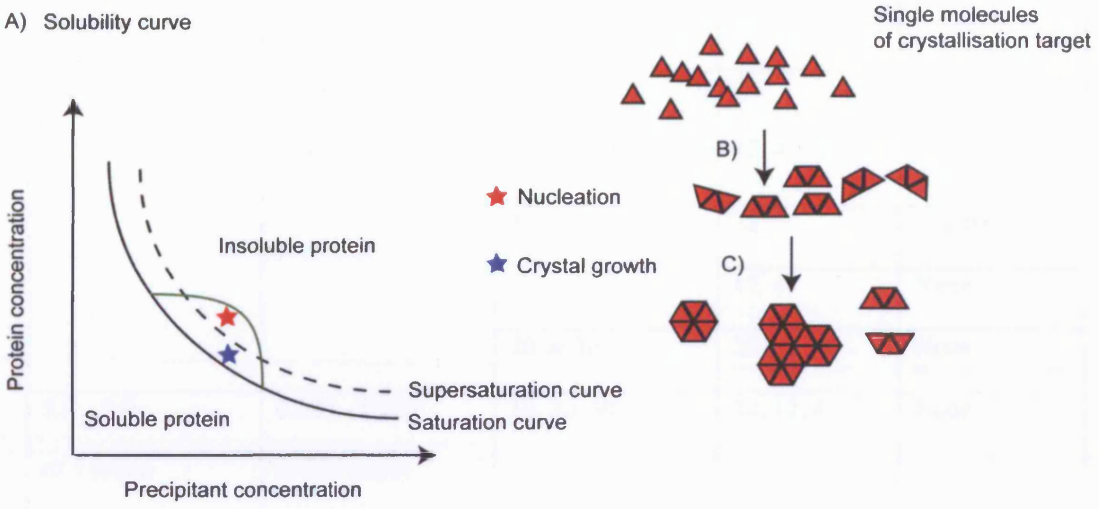

Figure 5.3. Nucleation and crystal growth

A) A phase diagram showing a hypothetical saturation curve, supersaturation curve and region of crystallisation of the protein target (green). The red star indicates the region in which B) nucleation occurs. The blue star indicates the region where C) crystal growth occurs. If the target is above the supersaturation curve then it can nucleate and crystal formation can begin. If the target is above the saturation curve, but is not supersaturated, then no nucleation will occur so no crystals will form.

The different conditions tested are summarised in Table 5.1. These conditions were based around the idea of trying to slow crystal growth. A reduction in temperature slows vapour diffusion between the hanging drop and the reservoir meaning the increase in protein concentration will be slower and therefore crystal growth might be slower. A reduction in the amount of PEG, or addition of glycerol, should make the *dr*RecD2/DNA S1 complex more soluble and thereby slow down crystal growth.

Table 5.1. Conditions tested to grow single *dr*RecD2/DNA S1 crystals

All conditions contained 0.1M MES pH 6.5.

Precipitant **PEG 8K (%w/v)**	**Salt** **Calcium acetate (M)**	**Additive** **Glycerol (% v/v)**	**Temperature (°C)**	**Crystal type**
8.5 – 5.5 (0.5 steps)	0.125 – 0.225 (0.025 steps)	0	22	Clusters
			12, 4	None
		10	22	Clusters
			12, 4	None
		20 or 30	22	None
5.0 – 3.0 (0.5 steps)	0.125 – 0.225 (0.025 steps)	10, 20, 30	22, 12, 4	None

No discrete crystals were grown using the above conditions. Where clusters grew, they did so within one week. Trays were checked weekly for two months. The thickest needles were observed in those crystal clusters grown in the presence of 0.175 M calcium acetate.

It appeared that lowering the precipitant (PEG 8K) concentration inhibited the spontaneous nucleation of the crystals, as did reducing the temperature or adding glycerol. Micro-seeding into these conditions would supply nucleated seeds, therefore by-passing the need for spontaneous nucleation. The crystal growth of the seeds might still occur, but at a slower rate and hence produce single crystals. Micro seeding was carried out into the conditions summarised in Table 5.2.

Table 5.2. Conditions tested in order to grow single *dr*RecD2/DNA S1 crystals by micro-seeding

All conditions contained 0.1M MES pH 6.5

Precipitant PEG 8K (% w/v)	**Salt Calcium acetate (M)**	**Additive Glycerol (% v/v)**	**Temperature (°C)**	**Crystal type**
5.0 – 1.0 (1.0 steps)	0.175	0, 10	22	Discrete and clustered
			12, 4	None
0	0.175	0	22	Discrete
			12, 4	None
		10	22, 12, 4	None

Discrete crystals were grown when the concentration of PEG 8K was reduced. As the concentration reduced from 5 % to 0 % (w/v) discrete crystals were favoured. When no PEG 8K was present only discrete crystals grew. The biggest crystal (100 x 200 x 2 μM) grew in 0.1M MES pH 6.5 and 0.125 - 0.175 M calcium acetate (see Figure 5.4). All the discrete crystals were very thin plates, the thinnest dimension being 1-2 μM.

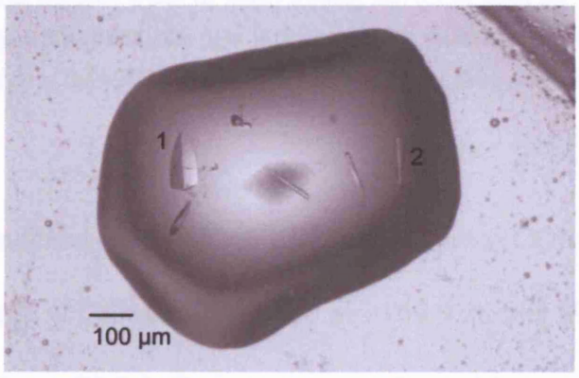

Figure 5.4. Discrete crystals of *dr*RecD2/DNA S1 complex

A picture of discrete crystals grown by microseeding into 0.1M MES pH 6.5, 0.175 M calcium acetate. Crystal 1 is lying with the thin edge across the plane of the picture. Crystal 2 is lying with the thin edge at right angles to the plane of the picture.

An attempt to capillary-mount these crystals in order to see if they would diffract at room temperature was made, but the crystals were so thin it proved very difficult not to damage them in the process. Instead, a screen for suitable cryoprotectant conditions was carried out. The crystals dissolve in 20-30% glycerol, ethylene glycol, PEG 400, and MPD. However, the crystals could tolerate MPD if the concentration of MPD was increased from 5 % to 20 % in quick 5% steps. The frozen crystals diffracted to around 5Å in-house.

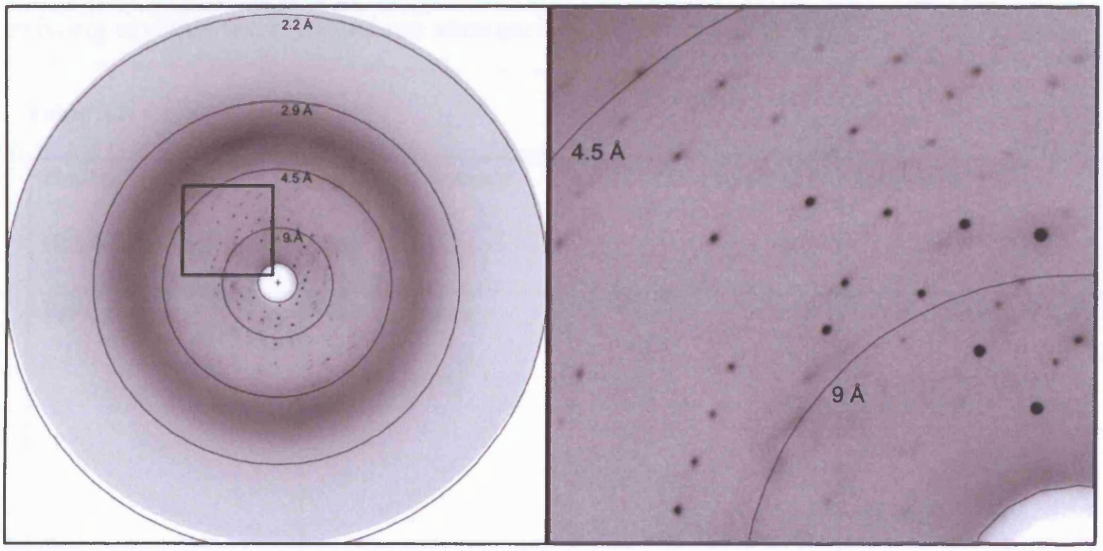

Figure 5.5. X-ray diffraction pattern of a discrete crystal of *dr*RecD2/DNA substrate 1 complex

X-ray diffraction pattern of a crystal grown by micro seeding into 0.1 M MES pH 6.5, 175 mM calcium acetate and cryoprotected in 20% MPD. The right-side panel shows the overall diffraction pattern and the square box is zoomed in on in the left-side panel. Resolution rings are shown with the resolution of each ring indicated.

The crystals were indexed using MOSFLM (CCP4 suite (CCP4 1994)). The space group was C2 and unit cell dimensions were a = 169.9 Å, b = 59.6 Å, c = 91.6 Å, α = 90.0 °, β = 98.6 ° γ = 90.0 °. Datasets were collected but the crystals were very anisotropic diffracting to around 4 Å in one direction while diffracting to 6 or 7 Å in the other direction. The resolution of a complete processed dataset was 6-7 Å. The

crystals were also highly mosaic, possibly as a result of damage during manipulation and freezing.

5.2.4 Optimisation of discrete crystals

It was reasoned that thicker crystals could be less anisotropic and more tolerant to freezing and could, therefore, provide higher resolution diffraction. In an attempt to grow thicker crystals different buffers were tried, and the conditions into which the existing crystals were seeded are summarised in Table 5.3 below.

Table 5.3. Changes in buffer

Buffer (0.1M and pH 6.5)	**Calcium acetate (M)**	**PEG 8K (% w/v)**	**Crystals**
PIPES	0.125 – 0.225 (0.025 steps)	0	Well-defined, thick discrete crystals.
		4	Thick but rough crystals
Sodium cacodylate	0.125 – 0.225 (0.025 steps)	0	Discrete plates
		4	Clustered plates
Imizadole	0.125 – 0.225 (0.025 steps)	0	Thin needles
		4	Clustered needles
Bis Tris	0.125 – 0.225 (0.025 steps)	0	None
		4	None

Crystals grew with PIPES, sodium cacodylate and imizadole buffers when microseeding was used. The crystals with the sharpest edges and thickest dimensions were obtained from PIPES buffer (see Figure 5.6).

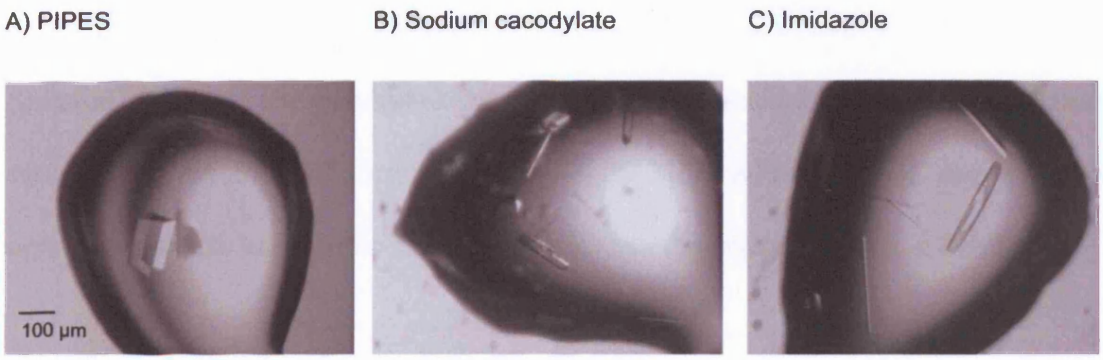

Figure 5.6. Crystals grown in PIPES, sodium cacodylate or imizadole

Crystals grown in 0.1 M A) PIPES, B) sodium cacodylate or C) imizadole, with 0.175 M calcium acetate using microseeding.

The crystals grown in PIPES, sodium cacodylate and imizadole were thicker than those grown in MES and could be capillary mounted to test for diffraction at room temperature. The crystals grown in sodium cacodylate and imizadole did not diffract. However, the crystals grown in PIPES diffracted to around 4-5 Å at room temperature in-house (see Figure 5.7). The intensity of the X-ray diffraction spots from crystals grown in PIPES were much greater than those from the crystals grown in MES.

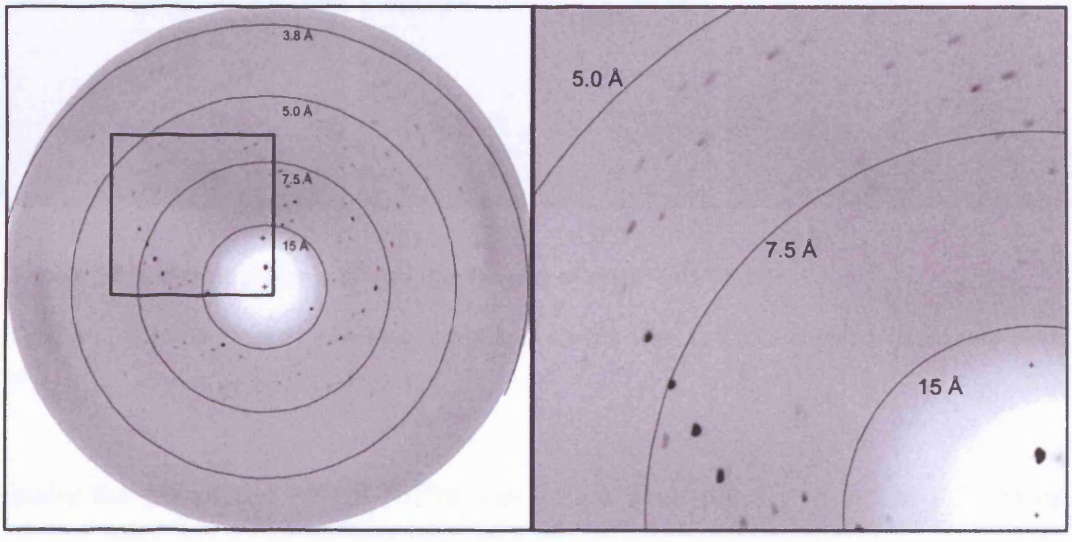

Figure 5.7. Room temperature X-ray diffraction of a discrete crystal of *dr*RecD2/DNA S1 complex

Crystals were grown in 0.1 M PIPES pH 6.5, 0.175 M calcium acetate. The right-side panel shows the overall diffraction pattern and the square box is zoomed in on in the left-side panel. Resolution rings are shown with the resolution of each ring indicated.

Suitable cryoprotectant conditions were screened, as described earlier, with the same results. The crystals were cryoprotected 20 % MPD added in 5 % steps. These crystals were less anisotropic and less mosaic than the crystals grown in MES.

Different salts of similar ionic strength to calcium acetate and containing calcium or acetate were tested including zinc acetate, magnesium acetate, sodium acetate, cobalt acetate, and calcium chloride. However, crystals did not grow in these salts with the MES or PIPES buffer whether or not microseeding was used. The length of the ssDNA tail was also varied from 6 to 12 nucleotides and crystals could be grown when longer tails of 8, 9 and 10 nucleotides were used, with microseeding using crystals grown in PIPES as the seeding stock. The thickest and largest crystals grew using the original DNA S1 with a 7 nucleotide tail, when longer tails were used the crystals were thinner and smaller (see Figure 5.8).

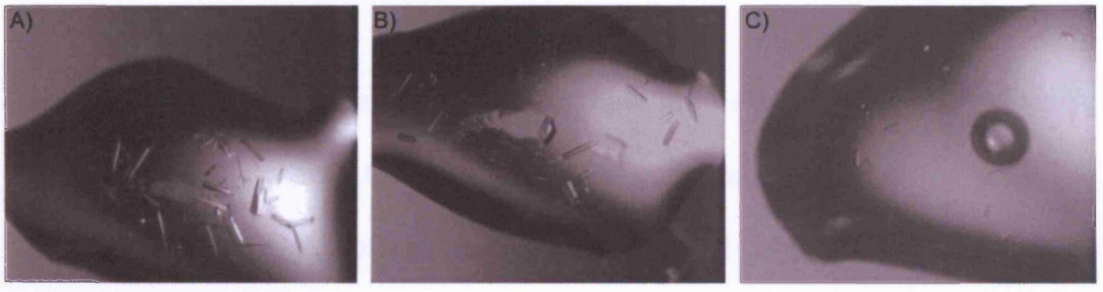

Figure 5.8. Crystals grown with varying lengths of the 5' ssDNA tail

Crystals of *dr*RecD2 DNA S1 complexes grown from seeding with A) 8 B) 9 and C) 10 nucleotide 5'-tails

Finally the pH of the PIPES buffer was varied from pH 6.1 to 7.5 in 0.2 pH unit steps. Crystals only grew when seeded into solutions buffered at pH 6.3, 6.5 and 6.7, with 6.5 yielding the biggest crystals.

The resulting crystals were inconsistent, with one in ten diffracting. Generally, bigger crystals diffracted to higher resolution and were more stable in the cryo-protectant used for freezing. The use of synchrotron facilities allowed rapid screening of, and data-collection from, smaller crystals.

5.2.5 Data collection

This section describes the collection of datasets that were used to try and solve the structure of the *dr*RecD2/DNA substrate 1 complex. All data collections were done at 100 °K after freezing the crystals.

The c axis of the unit cell sometimes doubled from around 90 Å (short c-axis) to around 180 Å (long c-axis). There was no obvious morphological difference between crystals with short and long c-axis. The data from crystals with the long c-axis could be processed by XDS.

Complete datasets from native crystals of both forms were collected. A native dataset from a crystal with the short c-axis was collected using beamline ID10.1 at the SRS, Daresbury. A native dataset from a crystal with the long c-axis was collected in-house.

5.2.5.1 Seleno-methionine *dr*RecD2

The *dr*RecD2 protein contains 15 methionines, excluding N-terminal methionine, and seleno-methionine *dr*RecD2 (se-met *dr*RecD2) was expressed in the *E.coli* B834 + strain, which cannot synthesise methionine. The *E.coli* were grown in rich medium with seleno-methionine replacing methionine. The se-met *dr*RecD2 was purified and crystallised using the same protocol as for the wild-type *dr*RecD2. The peak (pk),

remote (rm) and inflection (ip) wavelengths of the anomalous scattering were determined using a fluorescence wavelength scan of the crystal (see Figure 5.9).

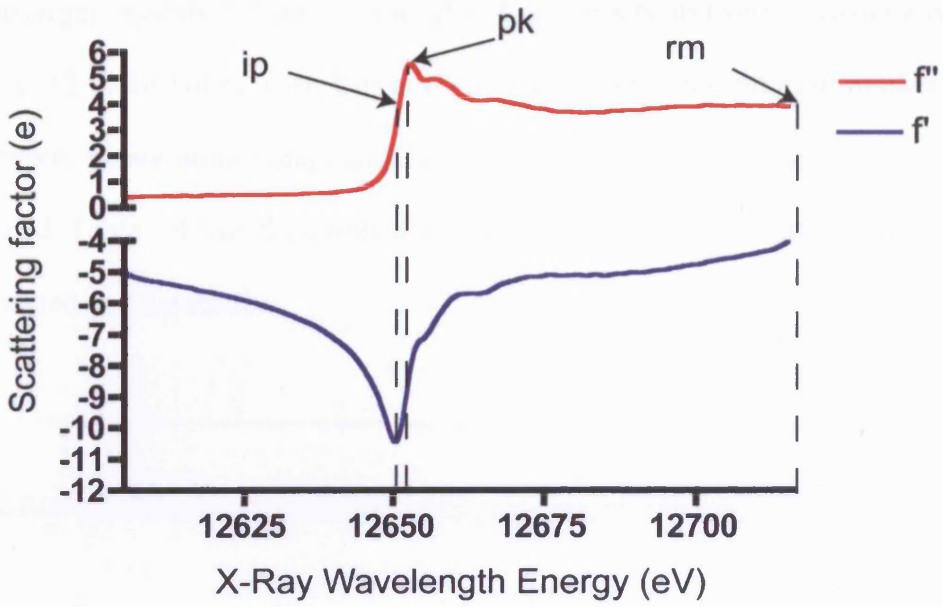

Figure 5.9. Fluorescence wavelength scan of seleno-methionine *dr*RecD2/DNA S1 crystal

Experimental fluorescence wavelength scan of seleno-methionine *dr*RecD2/DNA S1 crystal using beamline ID 10.1 at the SRS. The X-ray energy at the peak (pk), inflection (ip) and remote (rm) parts of the scan are indicated.

Three datasets were collected from the same crystal (with the short c-axis) at the peak, inflection and remote X-ray wavelengths using beamline ID 10.1 at the SRS Daresbury. Three datasets were also collected from the same crystal (with the long c-axis) at the peak, inflection and remote X-ray wavelengths using beamline ID 23.1 at the ESRF, Grenoble.

5.2.5.2 Heavy atom derivatives

In order to soak crystals with heavy atom compounds, stable harvest conditions for consistency between native crystals need to be found.

The majority of crystals did not survive for more than 4-6 hours when harvested out of the growth drop into new mother liquor, they either cracked or dissolved.

Conditions could not be found where the majority of harvested crystals survived for long time periods when removed from the drop they had grown in. However, some of the larger crystals did survive for up to 4 or 5 hours and very occasionally for as long as 12 to 18 hours when harvested into 1.1 times concentration mother liquor. Therefore, heavy atom compounds that do not require very long soak times were screened. Table 5.4 lists the compounds screened, the mode of action, reason for use, soaks tried and the results.

Table 5.4. Summary of heavy atom screening

Heavy atom compound	Formula	Mode of action	Reason for use	Soak		Result of soaking
				Concentration (mM)	Time (Hours)	
Potassium tetrachloroplatinate (II)	K_2PtCl_4	Reacts with methionine and histidine residues.	Reaction will proceed at pH 6.5, no phosphate is present to displace the chloride.	10, 5, 2.5, 1, 0.5	6, 12, 18, 24	The majority of crystals did not survive the soak. The surviving crystals diffracted to around 7-8 Å, both in house and at the ESRF.
				10, 5, 2.5, 1	4, 2, 1	The majority of crystals did not survive. A dataset of 5 Å resolution was collected.
Potassium hexachloroplatinate (IV)	K_2PtCl_6	As above	As above	10, 5, 2.5, 1, 0.5	6, 12, 18, 24	The majority of crystals did not survive the soak. The surviving crystals diffracted to around 7-8 Å, both in house and at the ESRF.
Sodium tetrachloroaurate (III)	$NaAuCl_4$	As above	As above	10, 5, 2.5, 1, 0.5	6, 12, 18, 24	No crystals survived the soak.
Ethyl mercuric phosphate (EMP)	$(C_2H_5HgO)HPO_2$	Reacts with histidine and cysteine residues	Reaction will proceed at pH 6.5 and is highly reactive meaning short soak times.	10, 5, 2.5,	4, 2, 1	No crystals survived the soak.
				1	4,2,1	The majority of crystals did not survive the soak. A dataset of 3.5 Å resolution was collected.
Triethyl lead acetate	$(C_2H_5)_3Pb(C_2H_3O_2)$	As above	Reaction will proceed at pH 6.5	50 mM	12, 18, 24	The majority of crystals did not survive the soak. A dataset of 4.2 Å resolution was collected.

A dataset was collected in-house from a crystal (of the short c-axis) soaked in 50 mM triethyl lead acetate. A dataset was collected from a crystal (of the short c-axis) soaked in 1 mM ethyl mercuric phosphate for 2 hours using beamline ID 14.3 at the ESRF, Grenoble. The ethyl mercuric phosphate-soaked crystal gave the best quality diffraction of all the crystals in the project and diffracted to 3.5 Å resolution (see Figure 5.10).

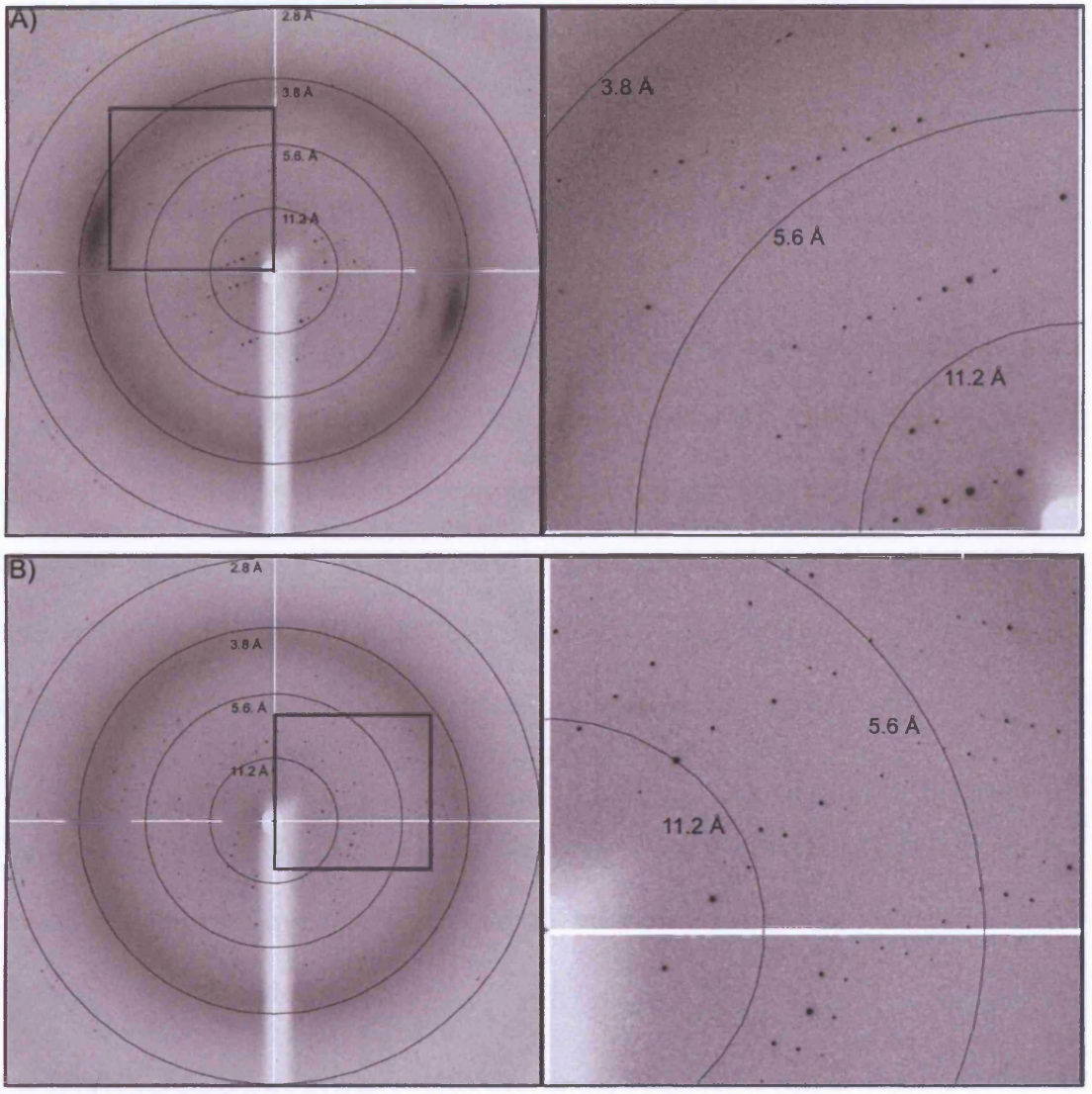

Figure 5.10. X-ray diffraction of a crystal of *dr*RecD2/DNA S1 complex soaked in ethyl mercuric phosphate

Crystals were grown in 0.1 M PIPES pH 6.5, 0.175 M calcium acetate and soaked in 1 mM ethyl mercuric phosphate for 2 hours. The right-side panel shows the overall diffraction pattern and the square box is zoomed in on in the left-side panel. Resolution rings are shown with the resolution of each ring indicated. A) Phi = 0 °. B) Phi = 90°.

5.2.5.3 Iodouridine DNA derivative

DNA can be modified to contain iodouridine bases inserted in place of thymines. All the thymines of the duplex portion of DNA S1 and alternate thymines of the pdT tail were substituted with 5-iodouridine to make iodouridine-substituted DNA S1 (iodo DNA S1).

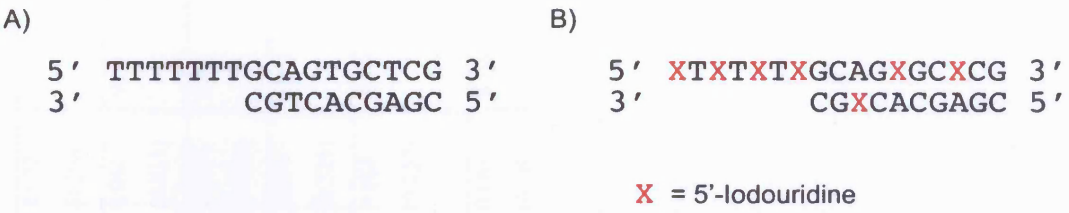

Figure 5.11. Iodouridine-substituted DNA S1

A) DNA S1 as used in the native crystals. B) Iodouridine-substituted DNA S1. X = 5' iodouridine.

The iodines in iodo DNA S1 would probably not provide sufficient phasing power for phase determination if used alone. They would, however, provide the register of the DNA and may help to further improve phases gained from other techniques. As the 5' tail comprises only thymines, trying to determine the position of the DNA bases in the electron density would be helped as every other thymine in the tail was replaced by the electron-dense iodouridine that could be easily identified.

The *dr*RecD2/iodo DNA S1 complex was crystallised as previously described. A dataset was collected in-house from a crystal with the long c-axis. No crystals with the short c-axis were identified.

5.2.6 Data processing

The processing statistics for the datasets of collected from crystals with the short c-axis are shown in Table 5. The processing statistics for the data sets collected from crystals with the long c-axis are shown in Table 5.6

Table 5.5. Summary of datasets from crystals with a short c-axis, processed in MOSFLM

Crystal		Unit cell (a, b, c (Å)) (α, β, γ (°))	Resolution (Å)	Mosaicity	Completeness (%)	Multiplicity	MeanI/sigmaI*	R factor*	Anomalous completeness (%)	Source	Wavelength
Native		172.5, 59.9, 90.9 90.0, 97.8, 90.0	4	0.86	99.3	3.6	7.0 (3.5)	0.072 (0.298)	-	SRS	0.9800
Se-met	pk	171.3, 59.9, 90.0 90.0, 98.1, 90.0	4	1.20	99.2	3.6	9.7 (3.4)	0.062 (0.211)	99.3	SRS	0.9800
	lp	172.1, 60.1, 90.4 90.0, 98.0, 90.0	4	1.12	99.2	3.5	8.2 (2.5)	0.070 (0.284)	99.1	SRS	0.98021
	rm	171.8, 60.0, 90.0 90.0 98.1, 90.0	4	1.18	99.3	3.6	8.0 (2.3)	0.079 (0.321)	99.3	SRS	0.97235
Ethyl mercury phosphate soak (EMP)		171.2, 59.6, 89.9 90.0, 99.0, 90.0	3.5	0.54	94.0	3.2	5.7 (2.3)	0.064 (0.257)	83.0	ESRF beam 14-3	0.93100
Lead Acetate soak		172.6, 59.6, 90.4 90.0, 98.7, 90.0	4.2	0.90	91.8	3.1	10.7 (3.0)	0.063 (0.263)	85.6	In house	1.5417

* Values in brackets indicate highest resolution shell.

Table 5.6. Summary of datasets from crystals with a long c-axis, processed in XDS

Crystal		Unit cell (a, b, c (Å)) (α, β, γ (°))	Resolution (Å)	Mosaicity (XDS)	Completeness (%)**	I/sigma*	R factor*	Source	Wavelength
Native		168.2, 58.6, 190.0 90.0, 93.5, 90.0	4.2	0.46	94.3	9.6 (2.8)	0.079 (0.350)	In-house	1.5417
Se-met	pk	168.3, 58.7, 187.8 90.0, 93.3, 90.0	4.0	0.39	91.1	10.1 (3.5)	0.056 (0.260)	ESRF 23-1	0.97915
	ip	168.0, 58.7, 188.1 90.0, 93.4, 90.0	4.0	0.39	94.1	11.8 (4.9)	0.069 (0.269)	ESRF 23-1	0.97940
	rm	167.7, 58.7, 188.0 90.0, 93.4, 90.0	4.0	0.38	94.2	10.3 (5.4)	0.078 (0.218)	ESRF 23-1	0.97565
Iodo DNA S1		164.8, 58.2, 186.4 90.0, 93.7, 90.0	4.2	0.37	97.5	10.3 (4.0)	0.088 (0.304)	In-house	1.5417

* Values in brackets indicate highest resolution shell.

** Includes anomalous completeness where appropriate.

5.2.7 Phase determination

5.2.7.1 Native crystals

Matthews coefficients of the native datasets were calculated using the Matthews coef programme (CCP4 suite (CCP4 1994)) (see Table 5.7).

Table 5.7. Matthews coefficient of native datasets

C-axis	Number of molecules per asymmetric unit	Matthews coefficient	Solvent (%)
Short c-axis	1	2.7	53
	2	1.3	7
Long c-axis	1	5.8	81
	2	2.9	63
	3	1.9	43
	4	1.4	24

The short c-axis crystal is likely to contain 1 *dr*RecD2/DNA S1 per asymmetric unit, and the long c-axis crystal is likely to contain 2 *dr*RecD2/DNA S1 per asymmetric unit.

5.2.7.2 Molecular replacement

It is very likely that *dr*RecD2 will possess the two RecA-like fold motor domains common to all SF1 helicases. Attempts to determine the phases by molecular replacement, using the programme Phaser (McCoy *et al.* 2005), were undertaken using various search models based on the structure of the *E.coli* RecD1 motor domains.

There is no sequence homology between the N-terminal domain 1 of *E.coli* RecD1 and the N-terminal region of *dr*RecD2, so this domain was not used in the search model; only the RecA like fold motor domains 1A and 2A were used. A search model was created from domains 1A and 2A by reducing the structure down to α-helices and β-strands only and removing any of these that may not be structurally conserved. This gave a core structure of the motor domains. Any amino acids that were not conserved between *dr*RecD2 and *E.coli* RecD1 were changed to alanine using the CHAINSAW programme (CCP4 suite (CCP4 1994)). Three search models were then made: (a) domains 1A and 2A fixed relative to each other as in the *E.coli* RecD1 structure, (b) domain 1A alone and (c) domain 2A alone. Each model was used with the native dataset and the higher resolution EMP dataset. Domain 1A alone (b) and domain 2A (c) alone were also used at the same time, to allow for any relative movement between the two in *dr*RecD2 compared to *E.coli* RecD1. No solutions could be identified in this way.

5.2.7.3 Heavy atom derivatives

The possible derivative datasets obtained from crystals soaked in heavy atoms were scaled against the native dataset using SCALEIT (CCP4 suite (CCP4 1994)). This provides various statistical outputs that allow the assessment of whether or not a heavy atom is present. The most important of these is the R factor for isomorphous differences. Using these data, it appeared that only the dataset from the ethyl mercuric phosphate-soaked crystal was likely to be a real derivative.

All of the datasets listed in Table 5., except that from lead acetate soaked crystal, were scaled together using SCALEIT (CCP4 suite (CCP4 1994)). A summary of the

R-factor for isomorphous differences (R_{iso}) between the datasets is shown in Table 5.8.

Table 5.8. Summary of R_{iso} between datasets

Data set	Native	Ethyl mercury phosphate	Se-met (pk)	Se-met (ip)
Native				
Ethyl mercury phosphate	0.209			
Se-met (pk)	0.110	0.198		
Se-met (ip)	0.106	0.202	0.066	
Se-met (rm)	0.122	0.128	0.066	0.075

SOLVE and auto.SHARP (Bricogne *et al.* 2003; Terwilliger and Berendzen 1999) were used to try and identify any heavy atom positions using the automatic techniques within the programmes. Corresponding peaks were looked for in difference and anomalous Patterson maps calculated using the FFT programme (CCP4 suite (CCP4 1994)). No heavy atom sites could be determined in this manner.

Another approach was taken, difference and anomalous Patterson maps were calculated at different resolution limits using the FFT programme (CCP4 suite (CCP4 1994)), and common peaks looked for. No large peaks (signal to noise ratio of greater than 3.5) were observed, but some peaks were common at all resolution limits. The positions were refined using SHARP, and the statistical outputs analysed. Two possible heavy atom sites were identified in this way, although the occupancy of both were very low. SHARP (Bricogne *et al.* 2003) also calculated an electron density map of using the phases calculated from the sites, but no coherent or protein-like electron density was observed. The phases calculated using these sites were then

used to try and find the selenium positions from the se-met *dr*RecD2 derivative using SOLVE and SHARP (Bricogne *et al.* 2003; Terwilliger and Berendzen 1999). No selenium positions were identified. One would expect to identify the positions of the seleniums if the mercury sites were correct. It was reasoned that the mercury sites were either not correct or that the non-isomorphous difference between the datasets were too large and masking any isomorphous differences from the presence of the heavy atoms.

The same approaches as above were used to try and identify selenium positions using isomorphous replacement with anomalous scattering (SIRAS), and also by multiple wavelength anomalous diffraction (MAD). No selenium positions were identified in this manner.

5.3 Discussion

No apo *dr*RecD2 crystals could be grown; this could be for any number of reasons. One of the reasons could be that the *dr*RecD2 was not pure enough and the impurities were interfering with the crystallisation process. This does not seem plausible because the same impurities would be present when *dr*RecD2/DNA S1 crystals were grown. Another reason may be structural non-uniformity between the *dr*RecD2 molecules. For example, the N-terminal amino acids 1 – 150 which may be a region of secondary structure of the protein may not be positioned in the same orientation in each *dr*RecD2 molecule. A uniformity maybe induced by the binding of the protein to DNA and a *dr*RecD2/DNA S1 complex did indeed crystallise. The crystals were optimised from clustered crystals to thin plates to thick crystals. PIPES was chosen as a buffer to try during optimisation as it is similar to MES, and PIPES produced thicker crystals than MES.

An experimental electron density map of the *dr*RecD2/DNA S1 complex was not obtained because the phases were not determined. The major problem when trying to determine the phases was the inconsistency between crystals. The unit cell of different crystals varied by between 0.5 and 2 Å in each axis, this could be due to several reasons. The crystals may be inherently different as they grow, which could be caused by differing purities of protein preparations or by differences in the microseeding stocks. The harvesting and manipulation of the crystals may be damaging. The cryoprotectant may also be damaging the crystals, indeed the crystals seem very sensitive to, and dissolve in, many different cryoprotectants. The freezing of the crystal may also be damaging.

The lack of consistent unit cell dimensions between crystals means the isomorphous differences due to the presence of a heavy atom would be lost in noise caused by the differences in the unit cell. This is not a problem when using MAD to try and determine the phases using datasets collected from a se-met *dr*RecD2 crystal as the three datasets at different wavelengths were collected from the same crystal. However, the 16 seleniums in a 86 KDa complex could not be located; the diffraction quality and resolution were probably not good enough.

The molecular replacement technique is not affected by the lack of consistent unit cells between crystals as only one dataset is used. There could be many reasons for the failure of molecular replacement to determine the phases in this case. The success of molecular replacement depends on the quality of the search model, i.e. how similar the search model structure is to the structure of the protein trying to be solved. The search model based on *E.coli* RecD1 may not be similar enough to the structure of *dr*RecD2, or may be too small in comparison with the size of *dr*RecD2 to provide any solutions.

Another major obstacle in determining the phases, which applies to all the strategies tried, comes from the fact that 3.5 – 4 Å resolution data may not good enough to solve the problem. Although many people recommend cutting the resolution of the data to 4 or 5 Å when trying to determine phases, this derives from data that extends to greater than 3 Å, meaning the quality of the data at 4 – 5 Å resolution is likely to be very good. In hindsight it may have been a better strategy to carry on optimising the crystals, or to try and find entirely new crystallisation conditions to obtain crystals that diffracted to 3 Å or better and had a consistent unit cell, rather than trying to solve the structure with the crystals as they were.

5.4 Future work

There are various ways forward with this project that could not be explored because of time constraints. The existing *dr*RecD2/DNA S1 complex crystals could be further optimised to gain better resolution and consistency of unit cells between crystals. Stable harvest conditions that allow for long soaks of the crystals without damaging them need to identified. The cryoprotection conditions could also be improved, more cryoprotectants could be screened and different techniques of freezing investigated. Further crystallisation trials of the *dr*RecD2/DNA S1 complex in the presence of AMPPNP could be carried out.

It would be worthwhile taking a step backwards and trying again to crystallise *dr*RecD2 in the absence of DNA. If this structure could be solved then it could be used for as model for molecular replacement in the case of the *dr*RecD2/DNA S1 complex crystals. The *dr*RecD2$^{\Delta NT}$ could also go undergo crystallisation trials in the presence and absence of ADP or AMPPNP. Indeed at the time of writing these recommendations have been followed and apo *dr*RecD2$^{\Delta NT}$ has been crystallised. The crystals grow in 100 mM Tris pH 8.5 and 13 – 18 % (v/v) ethanol and diffract to 3 Å in house (personal communication. Wigley, D). Work is currently under way screening for heavy atom derivates to solve the structure, and good progress toward determining the phases has been made. If the structure of apo *dr*RecD2$^{\Delta NT}$ is solved then it could be used as a search model to solve the structure of the *dr*RecD2/DNA S1 complex by molecular replacement.

The *dr*RecD2$^{\Delta NT}$ protein is missing amino acids 1-150 of *dr*RecD2. Further truncations of *dr*RecD2 could be made in order to create *dr*RecD2 mutants that

comprise only the core helicase domains. Limited proteolysis digests and N-terminal sequencing could indicate the best place to truncate *dr*RecD2. If these *dr*RecD2 truncation mutants could be crystallised and the structures solved they too would be very good search models to solve the structure of the *dr*RecD2/DNA S1 complex.

Another technique to crystallise *dr*RecD2 or the *dr*RecD2/DNA S1 complex is based upon the fact that both apo *dr*RecD2 and the *dr*RecD2/DNA S1 complex precipitate when ADP is added to them. As crystallisation is an ordered precipitation then it is possible that a controlled addition of ADP to either apo *dr*RecD2 or to the *dr*RecD2/DNA S1 complex may induce them to crystallise. A similar way of crystallising a protein, *Bacillus stearothermophilus* DNA Gyrase B, has been reported (Tsai 1996).

6 Bibliography

Abdel-Monem, M. and H. Hoffmann-Berling (1976). Enzymic unwinding of DNA. 1. Purification and characterization of a DNA-dependent ATPase from Escherichia coli. *European journal of biochemistry / FEBS* **65**(2): 431-440.

Alberg, A. J., A. P. Lam and K. J. Helzlsouer (1999). Epidemiology, prevention, and early detection of breast cancer. *Current opinion in oncology* **11**(6): 435-441.

Ali, J. A. and T. M. Lohman (1997). Kinetic measurement of the step size of DNA unwinding by Escherichia coli UvrD helicase. *Science (New York, N.Y* **275**(5298): 377-380.

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith and K. Struhl (1988). Current protocols in molecular biology. *Greene publishing Associates and Wiley-Interscience, NY.*

Bailey, S., W. K. Eliason and T. A. Steitz (2007). The crystal structure of the Thermus aquaticus DnaB helicase monomer. *Nucleic acids research.*

Baker, K. E. and R. Parker (2004). Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Current opinion in cell biology* **16**(3): 293-299.

Battista, J. R. (1997). Against all odds: the survival strategies of Deinococcus radiodurans. *Annual review of microbiology* **51**: 203-224.

Battista, J. R., A. M. Earl and M. J. Park (1999). Why is Deinococcus radiodurans so resistant to ionizing radiation? *Trends in microbiology* **7**(9): 362-365.

Bhattacharya, A., K. Czaplinski, P. Trifillis, F. He, A. Jacobson and S. W. Peltz (2000). Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. *RNA (New York, N.Y* **6**(9): 1226-1235.

Bianco, P. R., L. R. Brewer, M. Corzett, R. Balhorn, Y. Yeh, S. C. Kowalczykowski and R. J. Baskin (2001). Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. *Nature* **409**(6818): 374-378.

Bianco, P. R. and S. C. Kowalczykowski (2000). Translocation step size and mechanism of the RecBC DNA helicase. *Nature* **405**(6784): 368-372.

Bird, L. E., J. A. Brannigan, H. S. Subramanya and D. B. Wigley (1998). Characterisation of Bacillus stearothermophilus PcrA helicase: evidence against an active rolling mechanism. *Nucleic acids research* **26**(11): 2686-2693.

Bird, L. E., H. Pan, P. Soultanas and D. B. Wigley (2000). Mapping protein-protein interactions within a stable complex of DNA primase and DnaB helicase from Bacillus stearothermophilus. *Biochemistry* **39**(1): 171-182.

Bjornson, K. P., M. Amaratunga, K. J. Moore and T. M. Lohman (1994). Single-turnover kinetics of helicase-catalyzed DNA unwinding monitored continuously by fluorescence energy transfer. *Biochemistry* **33**(47): 14306-14316.

Boehmer, P. E. and P. T. Emmerson (1992). The RecB subunit of the Escherichia coli RecBCD enzyme couples ATP hydrolysis to DNA unwinding. *The Journal of biological chemistry* **267**(7): 4981-4987.

Boule, J. B. and V. A. Zakian (2006). Roles of Pif1-like helicases in the maintenance of genomic stability. *Nucleic acids research* **34**(15): 4147-4153.

Bowman, G. D., M. O'Donnell and J. Kuriyan (2004). Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature* **429**(6993): 724-730.

Brendza, K. M., W. Cheng, C. J. Fischer, M. A. Chesnik, A. Niedziela-Majka and T. M. Lohman (2005). Autoinhibition of Escherichia coli Rep monomer helicase activity by its 2B subdomain. *Proceedings of the National Academy of Sciences of the United States of America* **102**(29): 10076-10081.

Bricogne, G., C. Vonrhein, C. Flensburg, M. Schiltz and W. Paciorek (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta crystallographica* **59**(Pt 11): 2023-2030.

Bruck, I. and M. O'Donnell (2001). The ring-type polymerase sliding clamp family. *Genome Biol* **2**(1): REVIEWS3001.

Bullock, W. O., J. M. Fernandez and J. M. Short (1987). XL-1 Blue, a highly efficient plasmid transforming RecA *Escherichia coli* strain with beta-galactoside selection. *Biotechniques*(5): 376-379.

Byrd, A. K. and K. D. Raney (2004). Protein displacement by an assembly of helicase molecules aligned along single-stranded DNA. *Nature structural & molecular biology* **11**(6): 531-538.

Byrd, A. K. and K. D. Raney (2005). Increasing the length of the single-stranded overhang enhances unwinding of duplex DNA by bacteriophage T4 Dda helicase. *Biochemistry* **44**(39): 12990-12997.

Byrd, A. K. and K. D. Raney (2006). Displacement of a DNA binding protein by Dda helicase. *Nucleic acids research* **34**(10): 3020-3029.

Cann, I. K., S. Ishino, I. Hayashi, K. Komori, H. Toh, K. Morikawa and Y. Ishino (1999). Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea. *Journal of bacteriology* **181**(21): 6591-6599.

Cantor, S. B., D. W. Bell, S. Ganesan, E. M. Kass, R. Drapkin, S. Grossman, D. C. Wahrer, D. C. Sgroi, W. S. Lane, D. A. Haber, *et al.* (2001). BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell* **105**(1): 149-160.

Caruthers, J. M. and D. B. McKay (2002). Helicase structure and mechanism. *Current opinion in structural biology* **12**(1): 123-133.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta crystallographica* **50**(Pt 5): 760-763.

Chao, K. L. and T. M. Lohman (1991). DNA-induced dimerization of the Escherichia coli Rep helicase. *Journal of molecular biology* **221**(4): 1165-1181.

Chapados, B. R., D. J. Hosfield, S. Han, J. Qiu, B. Yelent, B. Shen and J. A. Tainer (2004). Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell* **116**(1): 39-50.

Chedin, F. and S. C. Kowalczykowski (2002). A novel family of regulated helicases/nucleases from Gram-positive bacteria: insights into the initiation of DNA recombination. *Molecular microbiology* **43**(4): 823-834.

Chen, H. W., B. Ruan, M. Yu, J. Wang and D. A. Julin (1997). The RecD subunit of the RecBCD enzyme from Escherichia coli is a single-stranded DNA-dependent ATPase. *The Journal of biological chemistry* **272**(15): 10072-10079.

Cheng, W., K. M. Brendza, G. H. Gauss, S. Korolev, G. Waksman and T. M. Lohman (2002). The 2B domain of the Escherichia coli Rep protein is not required for DNA helicase activity.

Proceedings of the National Academy of Sciences of the United States of America **99**(25): 16006-16011.

Cheng, W., J. Hsieh, K. M. Brendza and T. M. Lohman (2001). E. coli Rep oligomers are required to initiate DNA unwinding in vitro. *Journal of molecular biology* **310**(2): 327-350.

Cheng, Z., D. Muhlrad, M. K. Lim, R. Parker and H. Song (2007). Structural and functional insights into the human Upf1 helicase core. *The EMBO journal* **26**(1): 253-264.

Conti, E. and E. Izaurralde (2005). Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Current opinion in cell biology* **17**(3): 316-325.

Czaplinski, K., Y. Weng, K. W. Hagan and S. W. Peltz (1995). Purification and characterization of the Upf1 protein: a factor involved in translation and mRNA degradation. *RNA (New York, N.Y* **1**(6): 610-623.

Dawson, P. E., T. W. Muir, I. Clark-Lewis and S. B. Kent (1994). Synthesis of proteins by native chemical ligation. *Science (New York, N.Y* **266**(5186): 776-779.

Dillingham, M. S., P. Soultanas and D. B. Wigley (1999). Site-directed mutagenesis of motif III in PcrA helicase reveals a role in coupling ATP hydrolysis to strand separation. *Nucleic acids research* **27**(16): 3310-3317.

Dillingham, M. S., P. Soultanas, P. Wiley, M. R. Webb and D. B. Wigley (2001). Defining the roles of individual residues in the single-stranded DNA binding site of PcrA helicase. *Proceedings of the National Academy of Sciences of the United States of America* **98**(15): 8381-8387.

Dillingham, M. S., M. Spies and S. C. Kowalczykowski (2003). RecBCD enzyme is a bipolar DNA helicase. *Nature* **423**(6942): 893-897.

Dillingham, M. S., D. B. Wigley and M. R. Webb (2000). Demonstration of unidirectional single-stranded DNA translocation by PcrA helicase: measurement of step size and translocation speed. *Biochemistry* **39**(1): 205-212.

Dillingham, M. S., D. B. Wigley and M. R. Webb (2002). Direct measurement of single-stranded DNA translocation by PcrA helicase using the fluorescent base analogue 2-aminopurine. *Biochemistry* **41**(2): 643-651.

Dionne, I., R. K. Nookala, S. P. Jackson, A. J. Doherty and S. D. Bell (2003). A heterotrimeric PCNA in the hyperthermophilic archaeon Sulfolobus solfataricus. *Molecular cell* **11**(1): 275-282.

Donmez, I. and S. S. Patel (2006). Mechanisms of a ring shaped helicase. *Nucleic acids research* **34**(15): 4216-4224.

Enemark, E. J. and L. Joshua-Tor (2006). Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442**(7100): 270-275.

Erzberger, J. P. and J. M. Berger (2006). Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annual review of biophysics and biomolecular structure* **35**: 93-114.

Evans, T. C., Jr., J. Benner and M. Q. Xu (1998). Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* **7**(11): 2256-2264.

Evans, T. C., Jr., J. Benner and M. Q. Xu (1999a). The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins. *The Journal of biological chemistry* **274**(26): 18359-18363.

Evans, T. C., Jr., J. Benner and M. Q. Xu (1999b). The in vitro ligation of bacterially expressed proteins using an intein from Methanobacterium thermoautotrophicum. *The Journal of biological chemistry* **274**(7): 3923-3926.

Evans, T. C., Jr. and M. Q. Xu (1999). Intein-mediated protein ligation: harnessing nature's escape artists. *Biopolymers* **51**(5): 333-342.

Futreal, P. A., Q. Liu, D. Shattuck-Eidens, C. Cochran, K. Harshman, S. Tavtigian, L. M. Bennett, A. Haugen-Strano, J. Swensen, Y. Miki, *et al.* (1994). BRCA1 mutations in primary breast and ovarian carcinomas. *Science (New York, N.Y* **266**(5182): 120-122.

Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research* **31**(13): 3784-3788.

Gorbalenya, A. E. and E. V. Koonin (1991). Endonuclease (R) subunits of type-I and type-III restriction-modification enzymes contain a helicase-like domain. *FEBS letters* **291**(2): 277-281.

Gorbalenya, A. E. and E. V. Koonin (1993). Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.* **3**(3): 419-429.

Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko and V. M. Blinov (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic acids research* **17**(12): 4713-4730.

Gorbalenya, A. E., E. V. Koonin and Y. I. Wolf (1990). A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS letters* **262**(1): 145-148.

Gulbis, J. M., Z. Kelman, J. Hurwitz, M. O'Donnell and J. Kuriyan (1996). Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell* **87**(2): 297-306.

Gupta, R. and R. M. Brosh, Jr. (2007). DNA repair helicases as targets for anti-cancer therapy. *Current medicinal chemistry* **14**(5): 503-517.

Gutman, P. D., J. D. Carroll, C. I. Masters and K. W. Minton (1994). Sequencing, targeted mutagenesis and expression of a recA gene required for the extreme radioresistance of Deinococcus radiodurans. *Gene* **141**(1): 31-37.

Ha, T., I. Rasnik, W. Cheng, H. P. Babcock, G. H. Gauss, T. M. Lohman and S. Chu (2002). Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature* **419**(6907): 638-641.

Hall, M. C. and S. W. Matson (1999). Helicase motifs: the engine that powers DNA unwinding. *Molecular microbiology* **34**(5): 867-877.

Hickson, I. D., C. N. Robson, K. E. Atkinson, L. Hutton and P. T. Emmerson (1985). Reconstitution of RecBC DNase activity from purified Escherichia coli RecB and RecC proteins. *The Journal of biological chemistry* **260**(2): 1224-1229.

Ivessa, A. S., J. Q. Zhou, V. P. Schulz, E. K. Monson and V. A. Zakian (2002). Saccharomyces Rrm3p, a 5' to 3' DNA helicase that promotes replication fork progression through telomeric and subtelomeric DNA. *Genes & development* **16**(11): 1383-1396.

Iwai, H. and A. Pluckthun (1999). Circular beta-lactamase: stability enhancement by cyclizing the backbone. *FEBS letters* **459**(2): 166-172.

Jeruzalmi, D., M. O'Donnell and J. Kuriyan (2001a). Crystal structure of the processivity clamp loader gamma (gamma) complex of E. coli DNA polymerase III. *Cell* **106**(4): 429-441.

Jeruzalmi, D., O. Yurieva, Y. Zhao, M. Young, J. Stewart, M. Hingorani, M. O'Donnell and J. Kuriyan (2001b). Mechanism of processivity clamp opening by the delta subunit wrench of the clamp loader complex of E. coli DNA polymerase III. *Cell* **106**(4): 417-428.

Johnson, D. S., L. Bai, B. Y. Smith, S. S. Patel and M. D. Wang (2007). Single-molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase. *Cell* **129**(7): 1299-1309.

Johnson, R. E., S. Prakash and L. Prakash (1994). Yeast DNA repair protein RAD5 that promotes instability of simple repetitive sequences is a DNA-dependent ATPase. *The Journal of biological chemistry* **269**(45): 28259-28262.

Kim, J. I., A. K. Sharma, S. N. Abbott, E. A. Wood, D. W. Dwyer, A. Jambura, K. W. Minton, R. B. Inman, M. J. Daly and M. M. Cox (2002). RecA Protein from the extremely radioresistant bacterium Deinococcus radiodurans: expression, purification, and characterization. *Journal of bacteriology* **184**(6): 1649-1660.

Kim, J. L., K. A. Morgenstern, J. P. Griffith, M. D. Dwyer, J. A. Thomson, M. A. Murcko, C. Lin and P. R. Caron (1998). Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure* **6**(1): 89-100.

Kim, S. H., J. Smith, A. Claude and R. J. Lin (1992). The purified yeast pre-mRNA splicing factor PRP2 is an RNA-dependent NTPase. *The EMBO journal* **11**(6): 2319-2326.

Korolev, S., J. Hsieh, G. H. Gauss, T. M. Lohman and G. Waksman (1997). Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP. *Cell* **90**(4): 635-647.

Korolev, S., N. Yao, T. M. Lohman, P. C. Weber and G. Waksman (1998). Comparisons between the structures of HCV and Rep helicases reveal structural similarities between SF1 and SF2 super-families of helicases. *Protein Sci* **7**(3): 605-610.

Kumaraswamy, E. and R. Shiekhattar (2007). Activation of BRCA1/BRCA2-associated helicase BACH1 is required for timely progression through S Phase. *Molecular and cellular biology.*

Lahaye, A., S. Leterme and F. Foury (1993). PIF1 DNA helicase from Saccharomyces cerevisiae. Biochemical characterization of the enzyme. *The Journal of biological chemistry* **268**(35): 26155-26161.

Lee, J. Y. and W. Yang (2006). UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell* **127**(7): 1349-1360.

Lejeune, F. and L. E. Maquat (2005). Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Current opinion in cell biology* **17**(3): 309-315.

Levin, M. K., M. Gurjar and S. S. Patel (2005). A Brownian motor mechanism of translocation and strand separation by hepatitis C virus helicase. *Nature structural & molecular biology* **12**(5): 429-435.

Lohman, T. M. and K. P. Bjornson (1996). Mechanisms of helicase-catalyzed DNA unwinding. *Annu Rev Biochem* **65**: 169-214.

Lohman, T. M., K. Chao, J. M. Green, S. Sage and G. T. Runyon (1989). Large-scale purification and characterization of the Escherichia coli rep gene product. *The Journal of biological chemistry* **264**(17): 10139-10147.

Lusser, A. and J. T. Kadonaga (2003). Chromatin remodeling by ATP-dependent molecular machines. *Bioessays* **25**(12): 1192-1200.

Mackintosh, S. G. and K. D. Raney (2006). DNA unwinding and protein displacement by superfamily 1 and superfamily 2 helicases. *Nucleic acids research* **34**(15): 4154-4159.

Mahdi, A. A., G. S. Briggs, G. J. Sharples, Q. Wen and R. G. Lloyd (2003). A model for dsDNA translocation revealed by a structural motif common to RecG and Mfd proteins. *The EMBO journal* **22**(3): 724-734.

Mathys, S., T. C. Evans, I. C. Chute, H. Wu, S. Chong, J. Benner, X. Q. Liu and M. Q. Xu (1999). Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene* **231**(1-2): 1-13.

Matsumiya, S., Y. Ishino and K. Morikawa (2001). Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from Pyrococcus furiosus. *Protein Sci* **10**(1): 17-23.

McCoy, A. J., R. W. Grosse-Kunstleve, L. C. Storoni and R. J. Read (2005). Likelihood-enhanced fast translation functions. *Acta crystallographica* **61**(Pt 4): 458-464.

Michaelis, L. and M. L. Menten (1913). Die kinetik der invertinwirkung. *Biochem. Z* **49**: 333-369.

Miki, Y., J. Swensen, D. Shattuck-Eidens, P. A. Futreal, K. Harshman, S. Tavtigian, Q. Liu, C. Cochran, L. M. Bennett, W. Ding, *et al.* (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science (New York, N.Y* **266**(5182): 66-71.

Minton, K. W. (1994). DNA repair in the extremely radioresistant bacterium Deinococcus radiodurans. *Molecular microbiology* **13**(1): 9-15.

Moore, K. J. and T. M. Lohman (1994a). Kinetic mechanism of adenine nucleotide binding to and hydrolysis by the Escherichia coli Rep monomer. 1. Use of fluorescent nucleotide analogues. *Biochemistry* **33**(48): 14550-14564.

Moore, K. J. and T. M. Lohman (1994b). Kinetic mechanism of adenine nucleotide binding to and hydrolysis by the Escherichia coli Rep monomer. 2. Application of a kinetic competition approach. *Biochemistry* **33**(48): 14565-14578.

Morris, P. D., A. J. Tackett, K. Babb, B. Nanduri, C. Chick, J. Scott and K. D. Raney (2001). Evidence for a functional monomeric form of the bacteriophage T4 DdA helicase. Dda does not form stable oligomeric structures. *The Journal of biological chemistry* **276**(23): 19691-19698.

Muir, T. W., D. Sondhi and P. A. Cole (1998). Expressed protein ligation: a general method for protein engineering. *Proceedings of the National Academy of Sciences of the United States of America* **95**(12): 6705-6710.

Nanduri, B., A. K. Byrd, R. L. Eoff, A. J. Tackett and K. D. Raney (2002). Pre-steady-state DNA unwinding by bacteriophage T4 Dda helicase reveals a monomeric molecular motor. *Proceedings of the National Academy of Sciences of the United States of America* **99**(23): 14722-14727.

Nathanson, K. L., R. Wooster and B. L. Weber (2001). Breast cancer genetics: what we know and what we need. *Nature medicine* **7**(5): 552-556.

New England Biolabs IMPACT-TWIN instruction manual. Version 1.2.

Niedziela-Majka, A., M. A. Chesnik, E. J. Tomko and T. M. Lohman (2007). B. Stearothermophilus PCRA monomer is a single stranded DNA translocase but not a processive helicase in vitro. *The Journal of biological chemistry.*

Noirot-Gros, M. F., P. Soultanas, D. B. Wigley, S. D. Ehrlich, P. Noirot and M. A. Petit (2002). The beta-propeller protein YxaL increases the processivity of the PcrA helicase. *Mol Genet Genomics* **267**(3): 391-400.

Novy, R., D. Drott, K. Yaeger and R. Mierendorf (2001). Overcoming the codon bias of *E.coli* for enhanced protein expression *inNovations*(12): 1-3.

Pause, A. and N. Sonenberg (1992). Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *The EMBO journal* **11**(7): 2643-2654.

Petit, M. A., E. Dervyn, M. Rose, K. D. Entian, S. McGovern, S. D. Ehrlich and C. Bruand (1998). PcrA is an essential DNA helicase of Bacillus subtilis fulfilling functions both in repair and rolling-circle replication. *Molecular microbiology* **29**(1): 261-273.

Pullman, M. E., H. S. Penefsky, A. Datta and E. Racker (1960). Partial resolution of the enzymes catalyzing oxidative phosphorylation. I. Purification and properties of soluble dinitrophenol-stimulated adenosine triphosphatase. *The Journal of biological chemistry* **235**: 3322-3329.

Putnam, C. D., S. B. Clancy, H. Tsuruta, S. Gonzalez, J. G. Wetmur and J. A. Tainer (2001). Structure and mechanism of the RuvB Holliday junction branch migration motor. *Journal of molecular biology* **311**(2): 297-310.

Record, M. T., Jr., S. J. Mazur, P. Melancon, J. H. Roe, S. L. Shaner and L. Unger (1981). Double helical DNA: conformations, physical properties, and interactions with ligands. *Annu Rev Biochem* **50**: 997-1024.

Rocha, E. P., E. Cornet and B. Michel (2005). Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLoS genetics* **1**(2): e15.

Roman, L. J. and S. C. Kowalczykowski (1989a). Characterization of the adenosinetriphosphatase activity of the Escherichia coli RecBCD enzyme: relationship of ATP hydrolysis to the unwinding of duplex DNA. *Biochemistry* **28**(7): 2873-2881.

Roman, L. J. and S. C. Kowalczykowski (1989b). Characterization of the helicase activity of the Escherichia coli RecBCD enzyme using a novel helicase assay. *Biochemistry* **28**(7): 2863-2873.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). Molecular cloning: a laboratory manual, 2nd edn. *Cold Spring Habour NY.*

Scheffzek, K., M. R. Ahmadian, W. Kabsch, L. Wiesmuller, A. Lautwein, F. Schmitz and A. Wittinghofer (1997). The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science (New York, N.Y* **277**(5324): 333-338.

Servinsky, M. D. and D. A. Julin (2007). Effect of a recD mutation on DNA damage resistance and transformation in Deinococcus radiodurans. *Journal of bacteriology.*

Severinov, K. and T. W. Muir (1998). Expressed protein ligation, a novel method for studying protein-protein interactions in transcription. *The Journal of biological chemistry* **273**(26): 16205-16209.

Seybert, A., D. J. Scott, S. Scaife, M. R. Singleton and D. B. Wigley (2002). Biochemical characterisation of the clamp/clamp loader proteins from the euryarchaeon Archaeoglobus fulgidus. *Nucleic acids research* **30**(20): 4329-4338.

Seybert, A., M. R. Singleton, N. Cook, D. R. Hall and D. B. Wigley (2006). Communication between subunits within an archaeal clamp-loader complex. *The EMBO journal* **25**(10): 2209-2218.

Seybert, A. and D. B. Wigley (2004). Distinct roles for ATP binding and hydrolysis at individual subunits of an archaeal clamp loader. *The EMBO journal* **23**(6): 1360-1371.

Sikora, B., R. L. Eoff, S. W. Matson and K. D. Raney (2006). DNA unwinding by Escherichia coli DNA helicase I (TraI) provides evidence for a processive monomeric molecular motor. *The Journal of biological chemistry* **281**(47): 36110-36116.

Simmons, R. M. and T. L. Hill (1976). Definitions of free energy levels in biochemical reactions. *Nature* **263**(5578): 615-618.

Singleton, M. R., M. S. Dillingham, M. Gaudier, S. C. Kowalczykowski and D. B. Wigley (2004). Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* **432**(7014): 187-193.

Singleton, M. R., M. S. Dillingham and D. B. Wigley (2007). Structures and Mechanism of Helicases and Nucleic Acid Translocases. *Annu Rev Biochem.*

Singleton, M. R., M. R. Sawaya, T. Ellenberger and D. B. Wigley (2000). Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell* **101**(6): 589-600.

Singleton, M. R., S. Scaife and D. B. Wigley (2001). Structural analysis of DNA replication fork reversal by RecG. *Cell* **107**(1): 79-89.

Singleton, M. R. and D. B. Wigley (2002). Modularity and specialization in superfamily 1 and 2 helicases. *Journal of bacteriology* **184**(7): 1819-1826.

Skordalakes, E. and J. M. Berger (2003). Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* **114**(1): 135-146.

Soultanas, P., M. S. Dillingham, F. Papadopoulos, S. E. Phillips, C. D. Thomas and D. B. Wigley (1999). Plasmid replication initiator protein RepD increases the processivity of PcrA DNA helicase. *Nucleic acids research* **27**(6): 1421-1428.

Soultanas, P., M. S. Dillingham, P. Wiley, M. R. Webb and D. B. Wigley (2000). Uncoupling DNA translocation and helicase activity in PcrA: direct evidence for an active mechanism. *The EMBO journal* **19**(14): 3799-3810.

Spurling, T. L., R. L. Eoff and K. D. Raney (2006). Dda helicase unwinds a DNA-PNA chimeric substrate: evidence for an inchworm mechanism. *Bioorganic & medicinal chemistry letters* **16**(7): 1816-1820.

Stano, N. M., Y. J. Jeong, I. Donmez, P. Tummalapalli, M. K. Levin and S. S. Patel (2005). DNA synthesis provides the driving force to accelerate DNA unwinding by a helicase. *Nature* **435**(7040): 370-373.

Studier, F. W. and B. A. Moffatt (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology* **189**(1): 113-130.

Subramanya, H. S., L. E. Bird, J. A. Brannigan and D. B. Wigley (1996). Crystal structure of a DExx box DNA helicase. *Nature* **384**(6607): 379-383.

Tam, J. P., Y. A. Lu, C. F. Liu and J. Shao (1995). Peptide synthesis using unprotected peptides through orthogonal coupling methods. *Proceedings of the National Academy of Sciences of the United States of America* **92**(26): 12485-12489.

Tamura, J. K. and M. Gellert (1990). Characterization of the ATP binding site on Escherichia coli DNA gyrase. Affinity labeling of Lys-103 and Lys-110 of the B subunit by pyridoxal 5'-diphospho-5'-adenosine. *The Journal of biological chemistry* **265**(34): 21342-21349.

Tanner, N. K., O. Cordin, J. Banroques, M. Doere and P. Linder (2003). The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Molecular cell* **11**(1): 127-138.

Telenti, A., M. Southworth, F. Alcaide, S. Daugelat, W. R. Jacobs, Jr. and F. B. Perler (1997). The Mycobacterium xenopi GyrA protein splicing element: characterization of a minimal intein. *Journal of bacteriology* **179**(20): 6378-6382.

Terwilliger, T. C. and J. Berendzen (1999). Automated MAD and MIR structure solution. *Acta crystallographica* **55**(Pt 4): 849-861.

Tsai, F. T. F. (1996). Crystallographic Studies of DNA Gyrase B Protein. D. Phil Thesis. *Lincoln College, Oxford Univeristy.*

Tuteja, N. and R. Tuteja (2004). Prokaryotic and eukaryotic DNA helicases. Essential molecular motor proteins for cellular machinery. *European journal of biochemistry / FEBS* **271**(10): 1835-1848.

Usuki, F., A. Yamashita, I. Higuchi, T. Ohnishi, T. Shiraishi, M. Osame and S. Ohno (2004). Inhibition of nonsense-mediated mRNA decay rescues the phenotype in Ullrich's disease. *Annals of neurology* **55**(5): 740-744.

Usuki, F., A. Yamashita, I. Kashima, I. Higuchi, M. Osame and S. Ohno (2006). Specific inhibition of nonsense-mediated mRNA decay components, SMG-1 or Upf1, rescues the phenotype of Ullrich disease fibroblasts. *Mol Ther* **14**(3): 351-360.

Velankar, S. S., P. Soultanas, M. S. Dillingham, H. S. Subramanya and D. B. Wigley (1999). Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**(1): 75-84.

Walker, J. E., M. Saraste, M. J. Runswick and N. J. Gay (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO journal* **1**(8): 945-951.

Wang, J. and D. A. Julin (2004). DNA helicase activity of the RecD protein from Deinococcus radiodurans. *The Journal of biological chemistry* **279**(50): 52024-52032.

Watanabe, T., Y. Ito, T. Yamada, M. Hashimoto, S. Sekine and H. Tanaka (1994). The roles of the C-terminal domain and type III domains of chitinase A1 from Bacillus circulans WL-12 in chitin degradation. *Journal of bacteriology* **176**(15): 4465-4472.

Weng, Y., K. Czaplinski and S. W. Peltz (1996a). Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. *Molecular and cellular biology* **16**(10): 5477-5490.

Weng, Y., K. Czaplinski and S. W. Peltz (1996b). Identification and characterization of mutations in the UPF1 gene that affect nonsense suppression and the formation of the Upf protein complex but not mRNA turnover. *Molecular and cellular biology* **16**(10): 5491-5506.

White, O., J. A. Eisen, J. F. Heidelberg, E. K. Hickey, J. D. Peterson, R. J. Dodson, D. H. Haft, M. L. Gwinn, W. C. Nelson, D. L. Richardson, *et al.* (1999). Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1. *Science (New York, N.Y* **286**(5444): 1571-1577.

Wong, I. and T. M. Lohman (1992). Allosteric effects of nucleotide cofactors on Escherichia coli Rep helicase-DNA binding. *Science (New York, N.Y* **256**(5055): 350-355.

Wong, I. and T. M. Lohman (1993). A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proceedings of the National Academy of Sciences of the United States of America* **90**(12): 5428-5432.

Wu, H., M. Q. Xu and X. Q. Liu (1998). Protein trans-splicing and functional mini-inteins of a cyanobacterial dnaB intein. *Biochimica et biophysica acta* **1387**(1-2): 422-432.

Xi, X. G. (2007). Helicases as antiviral and anticancer drug targets. *Current medicinal chemistry* **14**(8): 883-915.

Xu, R., B. Ayers, D. Cowburn and T. W. Muir (1999). Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proceedings of the National Academy of Sciences of the United States of America* **96**(2): 388-393.

Yarranton, G. T., R. H. Das and M. L. Gefter (1979a). Enzyme-catalyzed DNA unwinding. A DNA-dependent ATPase from E. coli. *The Journal of biological chemistry* **254**(23): 11997-12001.

Yarranton, G. T., R. H. Das and M. L. Gefter (1979b). Enzyme-catalyzed DNA unwinding. Mechanism of action of helicase III. *The Journal of biological chemistry* **254**(23): 12002-12006.

Yarranton, G. T. and M. L. Gefter (1979). Enzyme-catalyzed DNA unwinding: studies on Escherichia coli rep protein. *Proceedings of the National Academy of Sciences of the United States of America* **76**(4): 1658-1662.

Ye, J., A. R. Osborne, M. Groll and T. A. Rapoport (2004). RecA-like motor ATPases--lessons from structures. *Biochimica et biophysica acta* **1659**(1): 1-18.

Zhang, D. H., B. Zhou, Y. Huang, L. X. Xu and J. Q. Zhou (2006a). The human Pif1 helicase, a potential Escherichia coli RecD homologue, inhibits telomerase activity. *Nucleic acids research* **34**(5): 1393-1404.

Zhang, X. D., S. X. Dou, P. Xie, J. S. Hu, P. Y. Wang and X. G. Xi (2006b). Escherichia coli RecQ is a rapid, efficient, and monomeric helicase. *The Journal of biological chemistry* **281**(18): 12655-12663.

Zhou, J. Q., H. Qi, V. P. Schulz, M. K. Mateyak, E. K. Monson and V. A. Zakian (2002). Schizosaccharomyces pombe pfh1+ encodes an essential 5' to 3' DNA helicase that is a member of the PIF1 subfamily of DNA helicases. *Molecular biology of the cell* **13**(6): 2180-2191.

Zhou, Q., X. Zhang, H. Xu, B. Xu and Y. Hua (2007). A new role of Deinococcus radiodurans RecD in antioxidant pathway. *FEMS microbiology letters* **271**(1): 118-125.

7 Appendix

7.1 Sequence comparison of RecD

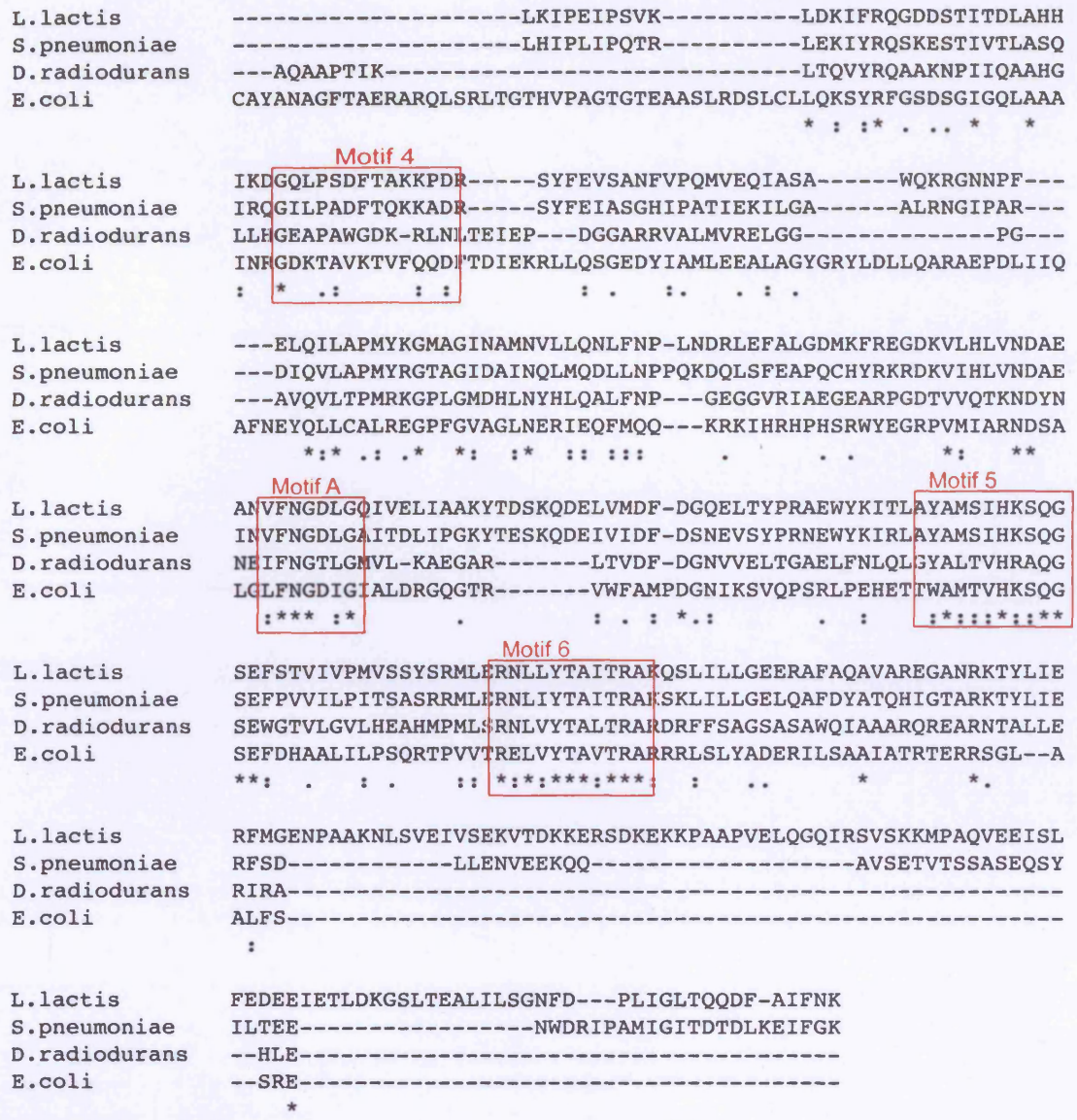

Figure 7.1. Sequence comparison of RecD

A sequence comparison of three type 2 RecD proteins, including *dr*RecD2, and the *E.coli* type 1 RecD protein. The helicase motifs are indicated as is the point at which *dr*RecD2 was truncated to create *dr*RecD2$^{\Delta NT}$. The sequence that corresponds to the molecular loop as described in chapter 4.5.5, and motif A as described in chapter 4.5.3 are also indicated.